2016

Substrate Recognition and Mechanistic Studies of Protein N-Terminal Methyltransferase 1

Yunfei Mao
Virginia Commonwealth University

SUBSTRATE RECOGNITION AND MECHANISTIC STUDIES OF PROTEIN N-TERMINAL METHYLTRANSFERASE 1

A dissertation submitted in partial fulfilment of the requirements for the degree of doctor of philosophy at Virginia Commonwealth University

by

YUNFEI MAO, BS

Advisor:
RONG HUANG, PHD
ASSISTANT PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY

Virginia Commonwealth University
Richmond, Virginia
June, 2016

Acknowledgements

This work is dedicated to all the people who have helped me, encouraged me, influenced me, and guided me. Without them, this accomplishment is impossible.

First of all, with all my gratitude, I would like to show my great appreciation to everybody from Dr. Rong Huang's lab. As Dr. Huang's first student, I have experienced different periods of time, but we always worked together, facing every problem we had encountered, and accomplished enormous amount of works. Dr. Huang has forgiven all those mistakes I had made and patiently guided me, and trained me from an ignorant undergraduate student to an accomplished enzymologist. Because of her, I may honourably graduate from this challenging program. And I thank Ms. Brianna Mackie my one year junior lab mate, as my friend and colleague. We have been through spectacular three years together. Here, I wish all my best to her last year in this program, and her future career!

I would like to thank, Drs. Yan Zhang, Glen Kellogg, Xianjun Fang, and Darrell Peterson as my committee members. Those valuable suggestions from them always kept me on the right track and allow me to move forward

I would like to thank our collaborators Drs. Jinrong Min and Cheng Dong for their dedication and collaboration in the structural work. And thank Drs. Glen Kellogg and Hardik Parikh for their collaboration in computational screening.

And I thank Drs. Martin Safo and Faik Musayev for their help in our protein crystallization. And finally, thank Dr. Peterson for our protein expression and purification, and thank his generosity to share his lab equipments and chemicals with us.

I would like to thank Dr. Chenxiao Da, as my first friend in the Department of Medicinal Chemistry, who helped me out in my first year here, and taught me tips of survival in graduate school. I would like to thank Drs. Kai Liu and Jeremy Chojnacki as my lunch companions in their last year in this department.

With my great honour, I would also like to thank my mentor from Waterford Institute of Technology (WIT), Dr. Hal Sherlock. He is a science maniac, a cantankerous chemist who is wild about electrical engineering. He helped, and guided me built up an ozone generator and an ozone sensor through recollecting disposed electrical parts. His passion about science had deeply affected me, and he showed me that science can be tremendously fun. In addition I would like to thank Dr. Patrick Duggan, the director of department of pharmaceutical science; he dedicated his time and energy in helping students with different problems, and made sure everybody could successfully earn their degrees.

Last of all, I would like to thank two of my dearest friends, Christopher Kent and Ken Mann, from Wexford of Ireland. Christopher, my first real friend from my Alma Mater (WIT), he is a person of virtue, a genuine man with integrity, and a solid insulator of any impurity. From him, I've learned how to give, and how to forgive. Ken Mann is a spectacular hiking expert, an inventor, and a father of a lovely teen-aging girl. Although we met during our senior year, the splendid moments we had together were more than

the addition to my first twenty-three years. He made me realize that life can be full of excitements if you pay attention and appreciate things around you.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

ATP	-	Adenosine triphosphate
CENP	-	Centromere protein
DDB2	-	DNA damaged-binding protein 2
GSH	-	Glutathione
HAT	-	Histone acetyltransferase
Hcy	-	Homocysteine
HDAC	-	Histone deacetylase
KAT	-	Lysine acetyltransferase
KDAC	-	Lysine deacetylase
MALDI	-	Matrix-assisted laser disportion/ionization
NRMT	-	N-terminal RCC1 methyltransferase
NTMT	-	N-terminal methyltransferase
PKMT	-	Protein Lysine methyltransferase
PRMT	-	Protein arginine methyltransferase
PTM	-	Post translational modification
RCC1	-	Regulator of chromatin condensation 1
Rb	-	Retinoblastoma
SAM	-	S-adenosyl methionine
SAH	-	S-adenosyl homocysteine
SAHH	-	S-adenosyl homocysteine hydrolase

Abstract

SUBSTRATE RECOGNITION AND MECHANISTIC STUDIES OF PROTEIN N-TERMINAL METHYLTRANSFERASE 1

By Yunfei Mao, B.S.

A dissertation submitted in partial fulfilment of the requirements for the degree of doctor of philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2016

Advisor:
Rong Huang, Ph.D.

The methylation at the α-N-terminal amines of proteins that start with a canonical motif X-P-K (X=A/P/S) has been a known modification for nearly four decades. In 2010, protein α-N-terminal methyltransferase 1 (NTMT1/NRMT1) was identified as the first enzyme responsible for this modification. NTMT2 was discovered as a second member belonging to this family, but it was reported as a mono-methylase. The identification of RCC1, retinoblastoma (Rb) protein, centromere protein-A/B (CENP-A/B), and DNA damaged-binding protein 2 (DDB2) as new NTMT1 substrates revealed NTMT1's biological significance in mitosis, cell-cycle regulation, centromere formation,

and damaged DNA repair, respectively. Although significant progress had been made, a clear understanding of how NTMT1 recognizes substrates remains to be determined. Also, there is no specific small molecule inhibitor for NTMT1.

To fill these gaps, we first established a fluorescence-based assay for kinetic characterization of NTMT1. Subsequently, ternary complex crystal structures of NTMT1 were obtained to illustrate the structural basis for enzyme-substrate interactions. The structures of the enzyme-substrate complex coupled with mutagenesis, binding, and enzymatic studies demonstrated the key elements involved in interaction with its substrates. In the meantime, we utilized computational studies and fluorescence assays for novel small molecule discovery. Lastly, we closely monitored the substrates' methylation progression by NTMT1 and NTMT2 in parallel using a MALDI-MS based assay.

Our results indicated that NTMT1 follows a Bi-Bi mechanism, and its methylation proceeds in a distributive pattern. Furthermore, NTMT1 was identified has broad substrate specificity beyond its canonical motif X-P-K (X=A/P/S), since X can be any amino acid except D/E and the third amino acids can also be R. We had also discovered an inhibitor that targets the substrate binding site of NTMT1 with IC_{50} = 7 μM. Lastly, our methylation progression studies has demonstrated that NTMT2 can also di-, tri-methylate certain substrates although its methylation rate is lower than NTMT1.

Overall, this project has laid the foundation for further investigation of N-terminal methylation in terms of functions, mechanisms, and inhibitor design.

1. Introduction

Post translational modification (PTM) is a biochemical process utilized by nature to alter the "chemical makeup" of various proteins and molecules.[1] In this process, substrates are covalently modified through "addition" or "removal" of different functional groups including acetylation, phosphorylation and methylation.[2-4] PTM is believed to be "nature's unique way to escape from genetic imprisonment."[2] So far, about 400 types of PTMs have been discovered, and more than 90,000 modifications have been identified. PTM has emerged as a major regulatory mechanism in different life forms through changes in subcellular localization, enzymatic activities, protein stability, and interactions with other proteins or molecules.[1]

PTMs are generally catalyzed by two types of enzymes. One type is referred as a "writer" such as the methyltransferases, acetyltransferases and kinases that add chemical groups to substrates. The other type is called an "eraser" such as the demethylase, deacetylase, and phosphatase that remove corresponding modifications from substrates.[1, 3] Generally, each PTM on each specific residue may represent a specific functional modification. In addition, combination of different types of PTM on the same substrate, leads to a diversified mechanism to regulate protein structures and functions.[3, 4] The histone code is a well-known example to illustrate such complexity of PTMs on its substrate, as the patterns of PTMs on the flexible histone tails regulate nucleosome assembly and chromatin structure.[5] Additionally, "reader" proteins that

recognize specific PTMs serve as a third dimension of regulation to guide and determine the biological outcome of certain PTMs.[5]

1.1 Acetylation

Acetylation refers to the covalent addition of an acetyl group to a protein.[6, 7] This modification was first identified on the epsilon amino group of a Lys residue of the histone protein in 1963.[5] Therefore, the first identified enzyme responsible for acetylation is named histone acetyltransferase (HAT).[8, 9] Recent studies discovered that acetylation can occur on many non-histone proteins such as tumor suppressor p53, transcriptional repressor protein YY1, high mobility group proteins, estrogen receptor α, hypoxia – inducible factor α and nuclear factor kappa-light-chain-enhancer of activated B cells.[10-14] Hence, HATs have been renamed as lysine acetyltransferases (KATs) to reflect their ability to install an acetyl group on the side chain of a lysine residue of non-histone proteins.[5, 15] On the other hand, lysine deacetylases (KDACs/HDACs) can remove the acetyl group from the side chain of lysine.[15] As interplays between KAT and KDAC regulate gene expression (**Figure1**), both KATs and KDACs are important epigenetic drug targets for various diseases including cancer, cardiovascular diseases, inflammation, and neurodegenerative diseases.[8]

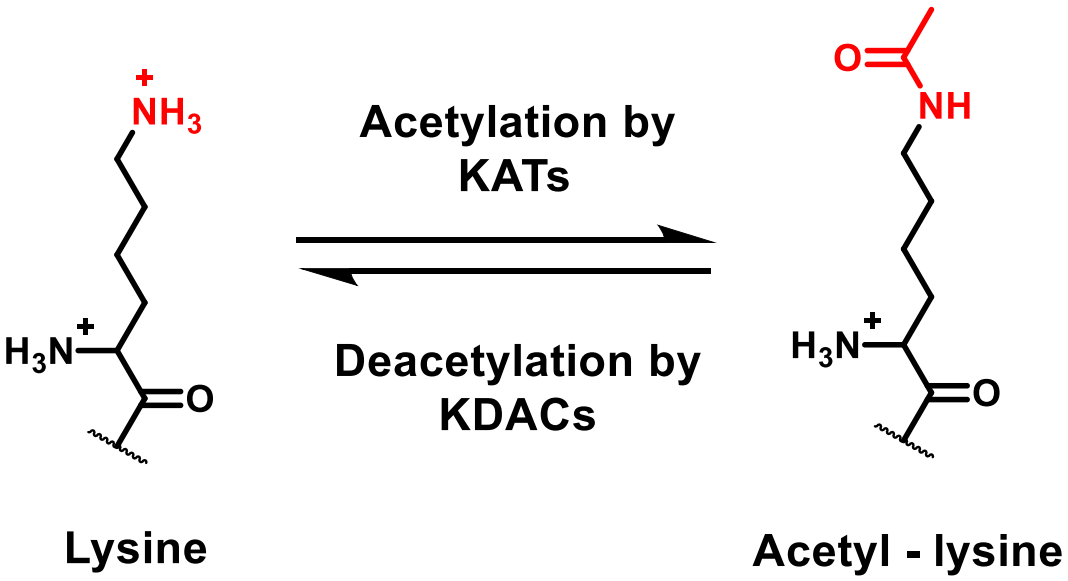


Figure 1. Protein acetylation and deacetylation.

Acetylation plays an important role in transcriptional activities, protein stability, and protein-protein interactions.[7, 16] Acetylation on the epsilon amino group of the lysine residue under physiological conditions modifies the overall electrostatic properties of the protein.[6, 9] Specifically, acetylation is one of the key epigenetic modifications that regulates chromatin structure because acetylated chromatin normally results in an open form and is thus accessible to various transcription factors (**Figure 2**).[5, 17, 18]

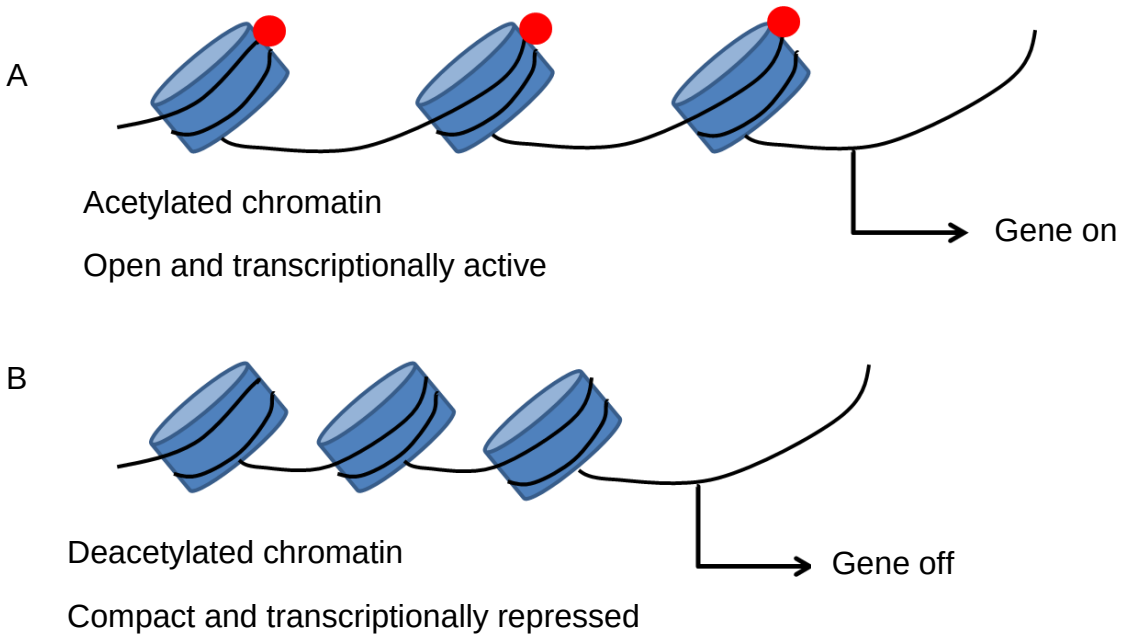


Figure 2. Acetylation and deacetylation of chromatin regulate gene expression. **A**. Acetylation (red sphere) on chromatins inhibits the folding of nucleosome arrays, which results in an open form to facilitate the access of transcription factors. **B**. Deacetylation of chromatins results in chromatin condensation, which in turn prohibits access of transcription factors.[5]

1.2 Phosphorylation

Phosphorylation refers to the addition of a phosphate group that is provided by a co-factor called adenosine triphosphate (ATP) to serine, threonine, and tyrosine residues on its target molecules.[19-25] Protein kinases are the enzymes responsible for phosphorylation modification.[26-28] Based on the residue of phosphorylated, kinases are divided into two categories: protein serine/threonine kinases and protein tyrosine kinases.[19, 29]

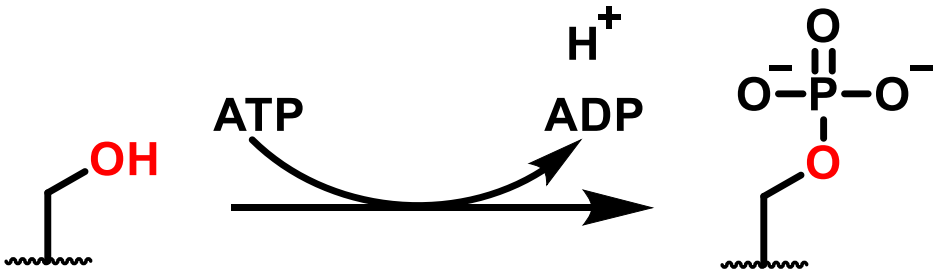


Figure 3. Phosphorylation modification.

Phosphorylation is intensively involved in regulating protein conformation activities and signal transduction.[19, 30] Under physiological conditions, the phosphoryl group is double negatively charged (**Figure 3**), and is frequently observed to have interactions with an Arg sidechain.[19] This interaction is known to stabilize the conformational state of a protein. Enzymes can be activated through phosphorylation through allosteric conformational changes.[19] One such example is glycogen phosphorylase, which exists in at least two functional states: the T (tense) state (a less active state), and the R (relaxed) state (a more actively state). The equilibrium between these two states is controlled by a phosphorylation switch.[23] Phosphorylation can also inhibit enzyme activity through steric blockage of substrate recognition site using phosphate group.[30] Additionally, in some cases, protein kinases may require phosphorylation at an allosteric site to induce a conformational change to create an active site for the subsequent phosphorylation.[31]

1.3 Methylation.

Methylation is a biochemical process of covalent addition of methyl groups to protein substrates.[32-34] The history of methylation can be traced back to early 1960s with the discovery of N-methyl-lysine in the flagella protein of *S.* typhimurium. During an investigation of the origin of N-methyl-lysine residues in histone, N-dimethy-lysine was identified in 1967 and followed by the identification of N-trimethyl-lysine in 1968.[32, 35, 36] Those discoveries initiated the pursuit of the enzyme responsible for such methylation reactions.[37] Interestingly, instead of discovering the protein responsible for lysine methylation, the first identified methylation protein was responsible for arginine methylation and is now known as protein arginine methyltransferse (PRMT).[32] Since the introduction of modern biochemical techniques in 1995, many various types of methylation proteins have been discovered and their important biological functions have been revealed in signal transduction, gene regulation, biosynthesis, and protein repair.[38, 39] Because of its significance, methylation has drawn attention and has become a rapidly expanding field.[32]

Methyltransferase is the enzyme responsible for the methylation modification.[33] It catalyzes the transfer of a methyl group from the substrate S-adenosylmethionine (SAM) to its substrates, which range from small molecules to proteins. SAM's primary role is donating a methyl group to different enzyme substrates. After the methyl group is transferred, SAM is converted to S-adenosylhomocysteine (SAH). So far, methyltransferases are thought to be the largest group of SAM-dependent enzymes (**Figure 4**).[33, 37]

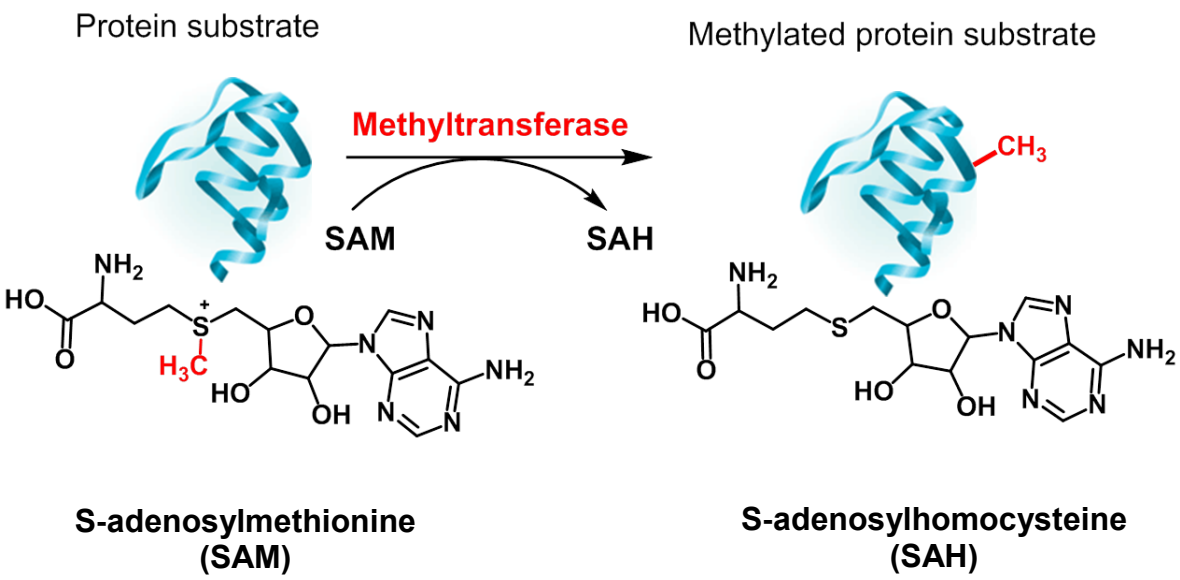


Figure 4. Protein methylation catalyzed by methyltransferse.[33]

1.3.1 Bisubstrate kinetic mechanisms

Protein methyltransferases belong to bisubstrate enzymes since they need to recognize both SAM and protein substrates. The kinetic mechanism of such bisubstrate enzyme is defined as the sequence of events in that its substrates are bound and products are released from the enzyme.[40-42] A bisubstrate enzyme has two possible mechanisms: a sequential (bi-bi) mechanism or a Ping-Pong mechanism.[40, 41, 43] As shown in **Figure 5**, in the sequential mechanism, both substrates “A” and “B” need to bind to the enzyme; “E” firstly to form a ternary complex “EAB” to trigger the catalytic activity of the enzyme. And both substrates will be converted into products “P” and “Q”, and subsequently released from the enzyme.

Figure 5. Sequential (Bi-Bi) mechanism **.**

In the Ping-Pong mechanism, as shown in **Figure 6**, one of the substrate “A” needs to bind to the enzyme “E” to “covalently” modify it, and it will be released as product “P”. And this modified enzyme “F” will subsequently bind to another substrate “B” to modify it and release it as product “Q”. After modification of substrates, “F” converts back to its original form “E”.

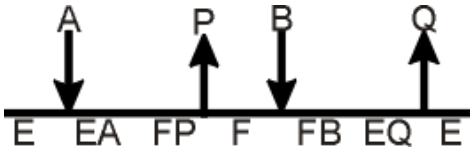


Figure 6. Ping-pong mechanism**.**

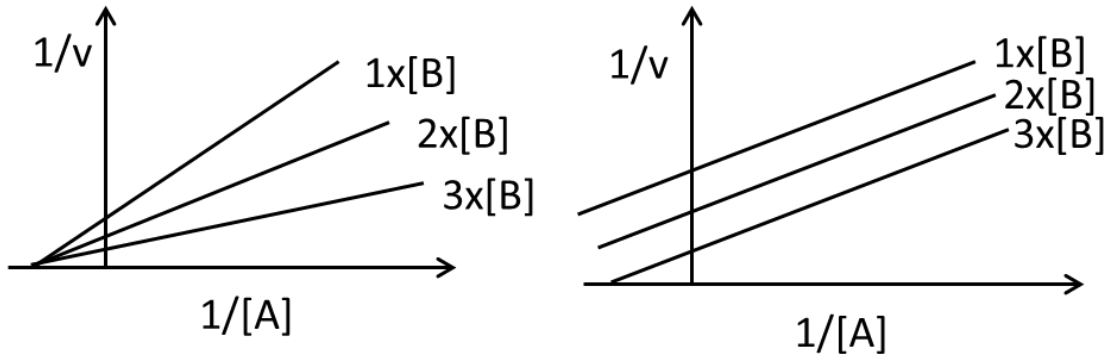
Understanding a kinetic mechanism is crucial for enzyme characterization, as well as inhibitor design. In order to determine the kinetic mechanism of a bisubstrate enzyme, one of the most common methodologies is the Lineweaver-Burk double reciprocal plot.[41, 42]

$$\frac{1}{V} = \frac{K_{\mathrm{m}}}{V_{\max}} \times \frac{1}{[S]} + \frac{1}{V_{\max}}$$

V: reaction rate
V_{max}: maximum enzyme reaction rate
[S]: Substrate concentration
K_m: substrate concentration that contributes to half V_{max}

Figure 7. Lineweaver-Burk equation.

This equation is derived from Michaelis-Menten kinetics model as indicated in **Figure 7**. Based on this equation, by plotting a graph of reaction rate reciprocal (1/V) *vs*. substrate concentration reciprocal (1/[S]), a linear curve can be obtained with a slope of K_m/V_{max} and its intersect at $1/V_{max}$ on y-axis. If an enzyme is adopting a sequential bi-bi mechanism, it should have an intersecting pattern as shown in **Figure 8**. Where, [A] and [B] represent corresponding substrate concentrations. Each curve represents different concentration of substrate B.



bi-bi mechanism **Ping-pong mechanism**

Figure 8. Lineweaver-Burk plot of Bi-Bi and Ping-Pong mechanism.

In contrast, if the enzyme is adopting Ping-Pong mechanism, a parallel pattern should be observed from the Lineweaver-Burk plot. [40-42]

1.3.2 Protein arginine methyltransferases (PRMTs).

PRMTs methylate the guanidinium nitrogen of specific arginine residues on histones (**Figure 9**).[44, 45] According to the types of methylated arginine products, PRMTs are divided into three subtypes. Type I PRMTs produce asymmetrically methylated di-methylarginine. And type II PRMTs produce symmetrically methylated di-methylarginine. Type III PRMTs catalyze arginine mono-methylation.[46] However, it is still unclear whether the mono-methylated product is the final product or an intermediate subject to further methylation.[47] So far, eight PRMTs have been identified in humans: PRMT1, PRMT3, PRMT4, PRMT6, PRMT8 belong to type I. And PRMT5, PRMT7 belong to type II.[46, 48]

It is known that PRMTs are extensively involved in gene expression regulation.[44, 49, 50] Among PRMTs, PRMT1 is involved in mRNA biosynthesis and heterochromatin formation. It also affects the subcellular localization of a number of its substrates.[44] Another noticeable member in this family is PRMT5, which regulates cell cycle, transcription, differentiation, stem cells, spliceosome assembly, and so on. Because of this, PRMT5 is an intriguing target for the study of different biological mechanisms.[51]

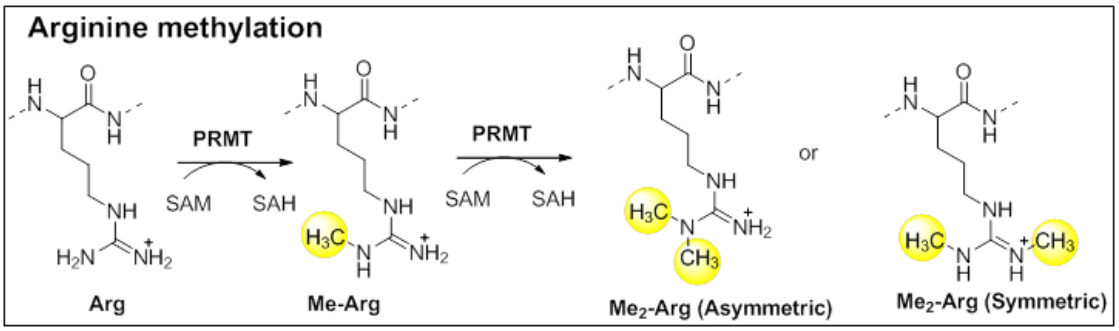


Figure 9. Protein arginine methylation.

Kinetic studies of PRMT1, PRMT5, and PRMT6 demonstrated that they all follow a Bi-Bi mechanism, which indicates the binding of both substrates to form a ternary complex.[42, 43, 52] The gel-based activity methods were utilized in those studies, which utilizing ^{14}C-labled SAM as the methyl donor. The incorporation of ^{14}C-labled methyl group to the arginine residue of substrates is monitored by phosphorimager to determine the rate of methylation.[42, 43, 52]

1.3.3 Protein lysine methyltransferases (PKMTs).

Site specific methylation of lysine on histone is regulated by a family of enzymes called PKMTs (**Figure 10**).[53] This family possesses a highly conserved SET domain and is well-known as SET-domain protein methyltransferases.[53-55] The SET domain is composed of approximately 130 residues; it was discovered as a conserved sequence in three *Drosophila melanogaster* proteins and was first characterized in 1998.[56, 57] So far, seven families of SET proteins are identified – SET 1, SET 2, SUV 39, EZ, RIZ, SMYD, and $SUV_{4\text{-}20}$.[57, 58] In addition to those family members described above, there are a few orphan members, such as SET 7/9 and SET 8.[59]

Although each PKMT has its specific roles, through methylating specific lysine with specific methylation states, PKMTs are extensively involved in epigenetic regulation of transcriptional activation, euchromatic/heterochromatic silencing, transcriptional elongation, and mitosis.[60-65] However, their roles are not confined to histone methylation; SET 7/9 was reported to methylate K189 of TAFF10, which is a general transcriptional factor. Additionally, SET 7/9 was identified to be involved in methylation of tumor suppressor p53, to increase its stability. Aberrant histone methylation is linked to developmental disorders and diseases.[59]

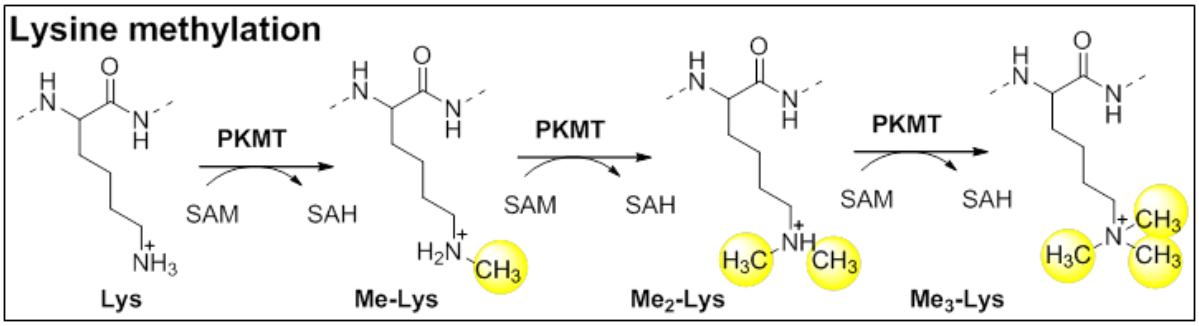


Figure 10. Protein lysine methylation.

1.4 Protein α-N-terminal methylation.

1.4.1 Discovery of α-N-terminal methylation

The first case of protein α-N-terminal methylation can be traced back to 1976 during a study of ribosomal subunits from *E. coli*.[66-69] It was identified that several ribosomal proteins; S11, L33 and L16, were methylated at their α-N-terminal amino groups.[67] Before 1987, different groups reported cases of N-terminal methylated proteins of varied species.[36, 70-72] Through sequence alignment of proteins that are subject to α-N-terminal methylation, several unique features of this modification were discovered. First, the first three amino acids at the N-termini of those proteins are relatively conserved, thus suggesting that they may serve as a recognition site for possibly enzymatic methylation. Second, the N-termini of prokaryotic and eukaryotic proteins possess different sequences. According to their methylation states and N-terminal specificity, prokaryotic protein can be divided into two classes. One class is those proteins subject to mono-methylation, which is comprised of ribosomal protein L16, chemotaxis CheZ protein, ribosomal protein L33, and the translational initiation factor IF-3. Those proteins were further divided into two subclasses based on their N-terminal sequences: L16 and CheZ possess glycine and proline at position 3 and 4, respectively. In contrast, IF-3 and L33 have either methionine or alanine at the N-terminus which is followed by lysine and glycine. In addition, there are two ribosomal protein subunits S11 and L11, they are also subject to N-terminal methylation. S11 has a unique N-terminal sequence of Ala-Lys-Ala. And L11 was the only identified prokaryotic protein that can be tri-methylated at that time, it has an N-terminal sequence of Ala-Lys-Lys (**Table 1**).[70]

Table 1. N-terminal sequence of prokaryotic proteins that are subject to N-terminal methylation.[70]

Protein	N-terminal sequence
L16	Met - Leu - Gln - Pro -
CheZ	Met - Met - Gln - Pro -
IF-3	Met - Lys - Gly - Gly -
L33	Ala - Lys - Gly - Ile -
S11	Ala - Lys - Ala - Pro -
L11	Ala - Lys - Lys - Val -

Compared to prokaryotic proteins that possess different N-terminal sequences and methylation states, all identified eukaryotic proteins that can be N-terminally methylated have a highly conserved N-terminal motif of Ala/Pro-Pro-Lys, and they all can be mono-, di-, tri-methylated.[70] At that time, the identified proteins that possess this motif included myosin light chain LC-1, histone H2B, and cytochrome c-557. As most of these proteins are part of macromolecular complexes, it is believed that N-terminal methylation regulates protein-protein interaction. Since all of these proteins contain a Pro-Lys motif at the second and third position of their N-termini, it was hypothesized that a single enzyme named “PK methyltransferase” was able to recognize the unique N-terminal motif to methylate those proteins. Unfortunately, due to the limited technologies and skills of that period, this “PK methyltransferase” was not identified.[70]

1.4.2 Discovery of protein N-terminal methyltransferase 1 & 2 (NTMT1/2)

In 2010 Webb et al. identified a protein named YBR261C/TAE1 that was responsible for N-terminal methylation of proteins through profiling of ribosomal proteins from yeast cells deficient in putative methyltransferases.[73] YBR261C/TAE1 is conserved across eukaryotes; its deletion strain showed abolished N-terminal methylation capability. Along with the discovery of YBR261C/TAE1, two human homologues, METTL11a and METTL11b were also identified.[73] Both YBR261C/TAE1 and METTL11a were demonstrated as active methyltransferases that recognize X-Pro-Lys N-terminal sequence, where X can be alanine, proline and serine. A further investigation of enzyme preference of the first residue of N-terminal sequence (X-Pro-Lys) suggests that those enzymes can recognize a variety of amino acids at the first position, and among which proline (Pro-Pro-Lys) had demonstrated the highest preference over other amino acids. Furthermore, through investigating the effects of substituting the second and third residues, it was found that both YBR261C/TAE1 and METTL11a prefer a Pro at position 2, and a Lys at position 3.[74] This result is in agreement with a previous study that all reported eukaryotic proteins subject to N-terminal methylation contain this motif.[73]

Interestingly, only three months after the first announcement of the identification of NTMT1 in 2010, Schanar-Tooley et al. had published the discovery of the first α-N-methyltransferase from Hela nuclear extracts and named it **N**-terminal **R**CC1 **m**ethyltransferase (NRMT).[75] NRMT is essentially the same enzyme discovered by Webb et al.[73] However, it is the first time that NRMT/NTMT1 was confirmed with

ELISA as a methyltransferase responsible for RCC1 N-methylation. Multi-spindle formation was observed during mitosis through knockdown of NRMT.[75]

Through docking studies and mutagenesis, only a few residues in the peptide substrate binding site were suggested to be essential for enzyme catalytic activity. For example, mutation of either residue N169 or D181 to lysine had abolished enzyme activity. Additionally, through an N-terminal sequence search of GenBank, the SET oncogene and Rb protein were identified as new substrates of NRMT/NTMT1. Both proteins were later confirmed biologically as authentic substrates of NTMT1. As mentioned before, the Rb protein is also known as the tumour suppressor, it is a regulator of cell cycle, and identification of this protein for N-terminal methylation was of great importance.[75-77]

A homologue named METTL11b was identified along with discovery of NTMT1.[73] However, a stable recombinant form of this protein was not obtained until 2013. It was found that the second N-terminal methyltransferase named NRMT2/NTMT2 has a similar localization as its homologue NTMT1, and recognizes the same N-terminal motif of X-P-K. However, it was reported to be mainly a mono-methylase as it only introduced one methyl group on its substrate.[78]

1.4.3 NRMT/NTMT1 substrate specificity

Since the discovery of N-terminal methylation, it was believed that the second residue of the NTMT1 substrates is always a highly conserved Pro (X-Pro-Lys).[70]

However in 2012, Petkowski et al. reported that the proline at the second position can be replaced by other residues: Ala, Glu, Met, Asn, Gln, Gly and Ser.[79]

In addition, Lys at position three had also been accepted as a highly conserved residue because replacing this lysine by glutamine was reported to result in diminished enzyme activity. Through testing the methylation of peptides with varied third residue, Petkowski et al. demonstrated that arginine can also fit into this position.[79] This was further confirmed when centromere protein A (CNEP-A) was identified as a new substrate of NTMT1 in 2013.[80, 81] It possesses a N-terminal sequence starting with Gly-Pro-Arg.[81-83] This study suggests that NTMT1 has broader substrate specificity than what was believed, and hence it was estimated that more than 300 proteins may be subject to N-terminal methylation based on this expanded substrate recognition.[81]

1.4.4 Identification of new substrates of NTMT1.

1.4.4.1 Regulator of chromatin condensation 1 (RCC1)

The knowledge of these N-terminal methylation modifications were largely uninvestigated until 2007, α-N-terminal methylation of a nucleotide - exchange factor named RCC1 was reported.[84] RCC1 is the only known guanine nucleotide exchange factor for Ran GTPase, which plays indispensable roles in nucleo-cytoplasmic transport, nuclear envelope assembly, and spindle formation in cell mitosis.[84-89] The association of RCC1 with chromatin through binding with histone H2A and/or H2B regulated by Ran is essential for RanGTP production (**Figure 11**).[84, 90]

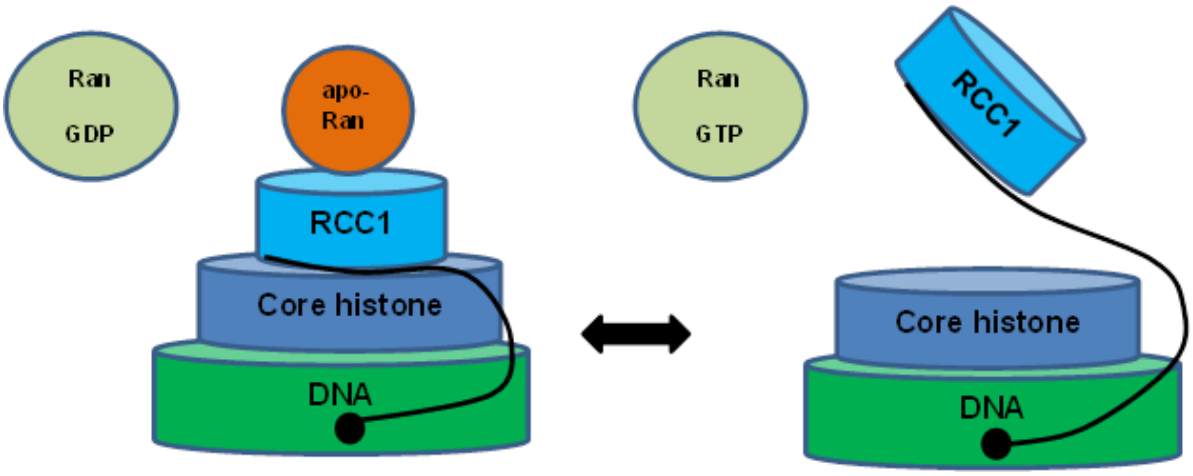


Figure 11. Diagram of RanGTP production.[75]

This is the first study that clearly demonstrated the function of N-terminal methylation and showed that methylated RCC1 N-terminus is essential for its association with chromatin, which in turn is crucial for cell mitosis. In order to examine the essentiality of RCC1 N-terminal methylation regarding its binding to chromatin, a series of RCC1 mutants were synthesized (APK-, PPK-, SPQ-, SPR-). Among them, mutant SPQ- showed abolished N-terminal methylation, and it was subsequently used for N-terminal methylation studies. Compared to wtRCC1, mutant RCC1 (SPQ-) with defected N-terminal methylation showed decreased binding efficacy to chromatin, which resulted in multi-spindle formation during mitosis. These results suggest the significance of RCC1 N-terminal methylation in cell mitosis.[84] The discovery of the significance of RCC1 N-terminal methylation rebooted an interest in studies about protein N-terminal methylation as well as providing the impetus for the discovery of NTMT1/NRMT1.

1.4.4.2 Centromere protein (CENP)

CENP is a chromatin region that serves as a spindle attachment point.[80-82] It plays essential roles in chromosome segregation during cell division.[90] CENP-B is a highly conserved centromere component that facilitates centromere formation in mammalian cells. It contains two important motifs: a DNA-binding motif at its N-terminus that binds specifically to a 17-bp DNA motif called CENP-B box within centromeric α-satellite DNA, and a dimerization domain at its C-terminus.[91, 92]

The N-terminus of CENP-B, which is the DNA-binding motif, possesses a Gly-Pro-Lys sequence.[80, 93-95] Hence it had been reasoned that CENP-B might be a new substrate of NTMT1, which was confirmed in 2013 by Dai et al. that N-terminal methylation of CENP-B by NTMT1 had strengthened its binding to centromeric DNA (**Figure 12**).[80]

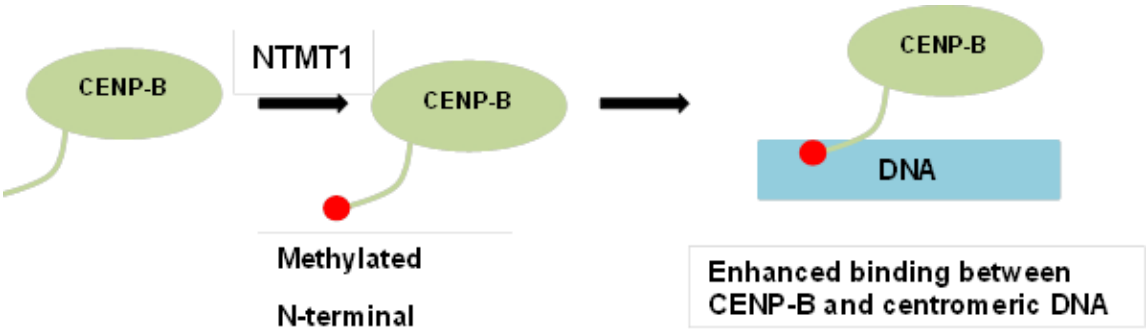


Figure 12. α-N-Terminal methylation of CENP-B.[80]

Meanwhile, CENP-A was also identified as a new substrate of NTMT1.[95] CENP-A is highly conserved among eukaryotes, and it is essential for the assembly of other centromeres. The function of N-terminal methylation of CENP-A remains elusive, but CENP-A contains a unique N-terminal motif of Gly-Pro-Arg.[91, 95] It had been previously known that NTMT1 can recognize synthetic peptides that have different amino acids at the first position and the third residue can be an Arg. However, CENP-A and CENP-B are two natural substrates that possess Gly at the first position and Arg at the third position (CENP-A).[91] Considering the significance and roles of those two proteins as histone variants in cell division, these discoveries again lit up the N-terminal methylation field.

1.4.4.3 Discovery of damaged DNA-binding protein 2 (DDB2).

The **UV**-**d**amaged **D**NA-**b**inding protein complex (UV-DDB) is one of two principal initiators of the nucleotide-excision repair (NER) pathway, which is responsible for the repair of different type of DNA damages.[96, 97] UV-DDB is a dimer complex composed of two subunits: a 127 kDa protein DDB1 and a 48 kDa protein DDB2, respectively. Through DDB2, UV-DDB binds specifically to the damaged site of DNA. Mutation of DDB2 was shown to cause cancer prone diseases such as xeroderma pigmentosum (XP).[98, 99]

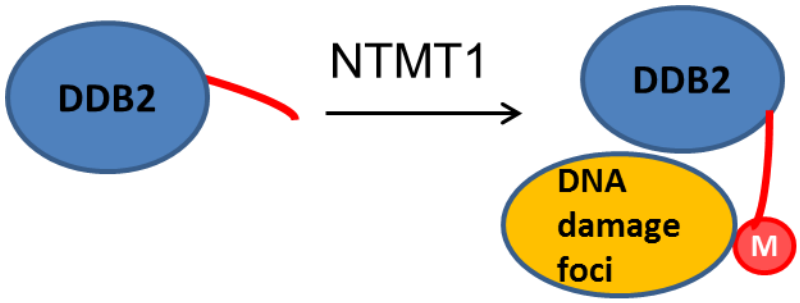


Figure 13. α-N-Terminal methylation of DDB2.[98]

DDB2 possesses an N-terminal motif of Ala-Pro-Lys, which suggested that DDB2 might also be a substrate of NTMT1 (**Figure 13**).[98] In 2014, through a LC-MS/MS based assay, Cai et al. demonstrated that DDB2 is in fact a substrate of NTMT1, and was mostly tri-methylated during the experiment. Moreover, DDB2 mutants with defective N-terminal methylation had demonstrated diminished nuclear localization and reduced recruitment to damaged DNA foci, suggested an indispensable role for DDB2 N-terminal methylation in UV-damaged DNA repair.[98]

1.4.5 NTMT1 bisubstrate inhibitors

The biological significance of NTMT1 in cell mitosis, damaged DNA repair, and its upregulation in cancers has also motivated studies of NTMT1 inhibitor design. Kinetic mechanism studies of PRMTs and PKMTs had illustrated that many members from this family are adopting a sequential (Bi-Bi) mechanism, which requires the formation of a ternary complex to initiate enzyme activity.[52] As a member of the methyltransferse family, we hypothesized that NTMT1 is likely to adopt this mechanism, and thus design and

synthesize a bisubstrate inhibitor by covalently linking a SAM analogue with a peptide substrate to mimic the ternary complex during enzyme catalysis. Such bisubstrate inhibitors could simultaneously inhibit both binding sites to provide potent and specific inhibitors.[100] The bisubstrate inhibitor was designed by using N-adenosyl-$_L$-methionine (NAM) to mimic SAM, which processes a more stable nitrogen instead of a active sulfonyl center. The N-terminal sequence derived from hRCC1 (SPKRIA) was used to mimic the peptide substrate of NTMT1. The NAM and SPKRIA are linked through a triazole linker based on our previous docking studies which shows that the sulfonyl group and α-amino group of the peptide is about 3.6 Å (**Figure 14**).[101]

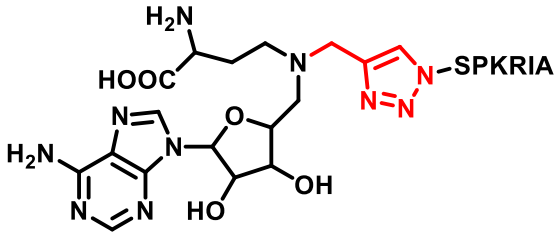


Figure 14. NTMT1 bisubstrate inhibitor.[101]

The designed bisubstrate inhibitor showed high inhibition potency (IC_{50} = 0.81 ± 0.13 μM), which also demonstrating selectivity: our results indicate that it has less than 15% inhibition effect on PRMT1 and less than 50% inhibition effect on G9a. This molecule is the first designed bisubstrate inhibitor of NTMT1.[101]

1.5 Specific aims of this study

It has been forty years since the first identification of protein α-N-terminal modification. In the meantime, the enzyme that is responsible for this type of modification was identified as NRMT/NTMT1. Since discovery, its substrate specificity has expanded from X-P-K (X = S, P, A) to X-P-K/R (X = S, P, A, G). Beside the natural substrates, it was demonstrated that NTMT1 can also methylate different synthetic peptides, e.g., where X can be most natural amino acids. Also, the conserved second residue P can be replaced by other residues, although this observations somewhat controversial.

Along with those discoveries, several new substrates of NTMT1 were also identified. Among them, RCC1, Rb protein, oncoprotein SET, CENP-A/B and DDB2, enhanced the significance of N-terminal methylation. Those substrates play critical roles in chromatin segregation, cell cycle regulation, centromere formation, and UV-damaged DNA repair, respectively.

However, studies of N-terminal methylation are still in an early stage. So far, most studies about this modification have been focusing on its substrates' function and substrate specificity. In order to understand N-terminal modification in a better perspective, it is important to obtain a comprehensive insight into the enzyme kinetics, substrate specificity, mechanisms, and inhibition of NTMT1. Hence, the specific aims of my graduate research are:

1. To develop biochemical assays to understand the kinetic mechanism of NTMT1.

2. To determine the mechanism of multi-step methylation progression by MALDI-MS

3. To elucidate the molecular basis of substrate recognition in a combination of crystal structures, site-directed mutagenesis, biochemical assays and binding studies to characterize the contribution of residues regarding substrate recognition.

4. To apply a combination of computational studies and biochemical assays to discover small molecule inhibitors for NTMT1.

5. To understand the product specificity of NTMT1/2 through MALDI-MS base methylation progression studies.

2. Results and Discussion

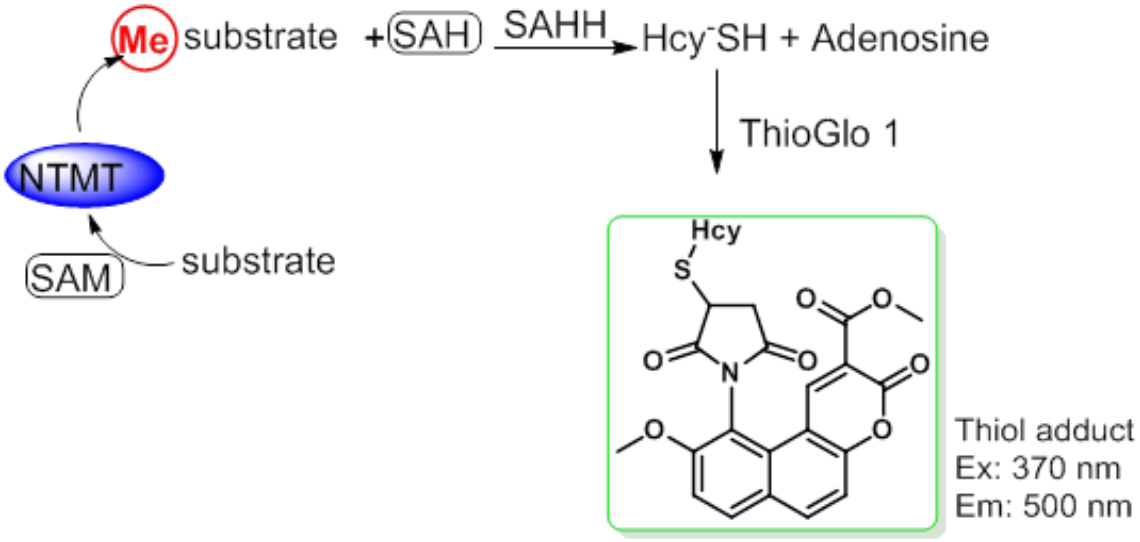
2.1 Fluorescence assay development

2.1.1 Design

First, we need to establish a convenient and sensitive method to quantify methylation in order to characterize the kinetic mechanism of NTMT1. We adapted a fluorescence-based SAH hydrolase (SAHH)-coupled assay (**Scheme 1)**, which monitors the conversion of SAM to SAH, using SAHH to catalyze the quantitative hydrolysis of SAH to adenosine and homocysteine (Hcy). Subsequently, the free thiol group of Hcy reacts with a sulfhydryl-sensitive fluorophore called ThioGlo1 to form an Hcy-Thioglo1 adduct. This adduct has a strong fluorescence at 500 nm when it is excited at 370 nm. The concentration of Hcy is subsequently determined by fluorescent intensity. In essence, the rate of SAH production is measured during enzyme catalysis reactions in this assay.

RCC1-12 peptide (SPKRIAKRRSPP) derived from the N-terminus of RCC1 and showed exothermic binding to NTMT1 with a K_d = 70 μM, was used this RCC1-12 as the NTMT1 substrate in the fluorescence-based assay. It was synthesized using standard Fmoc chemistry on an automated peptide synthesizer and purified by reverse phase. A correct mass of RCC1-12 (exact mass: 1390.8633) was confirmed via a MALDI-MS.

Scheme 1. Schematic diagram of fluorescence-based assay mechanism.



2.1.2 Assay optimization and validation

The correlation study between fluorescence intensity and free thiol concentration

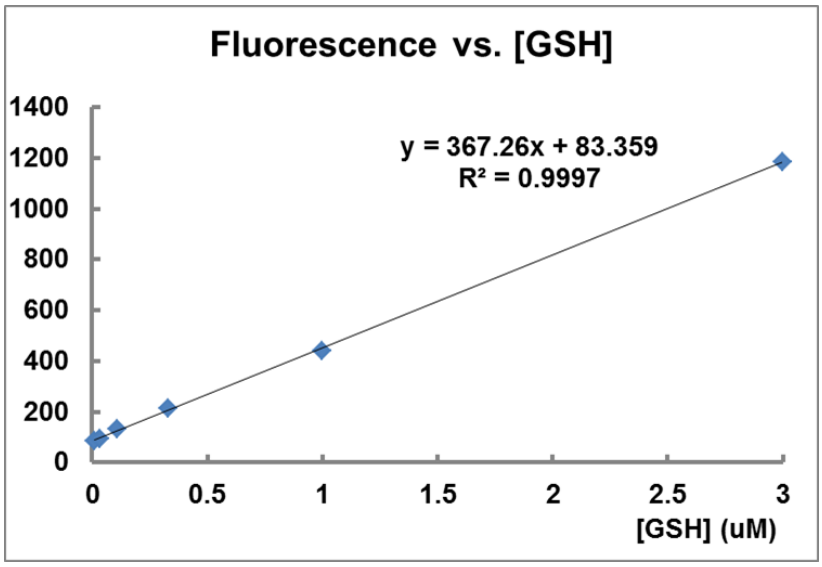


Figure 15. Standard curve of fluorescence intensity *vs.* [GSH].

Since the fluorescence intensity is our readout of the assay, the first thing was to ensure measured fluorescence intensity is proportional to product concentration. We used the commercially available glutathione (GSH) that contains a free thiol group to titrate the fluorescence intensity. The concentration range of GSH from 0 to 3 μM was examined in this study. We used excess amount of ThioGlo1 (15 μM) to ensure that ThioGlo1 is not a limiting factor. A standard curve of fluorescence intensity *vs.* GSH concentration was plotted in Microsoft Excel (**Figure 15)** and our results indicated that GSH concentration and measured fluorescence intensity is in linear relationship with R^2 = 0.9997.

Comparison of fluorescent dyes

To investigate the sensitivity of the fluorescent dye, we compared two commonly used fluorophores: ThioGlo1 and CPM. Both dyes are specific to free thiol groups. The difference is that CPM-thiol adduct generates a strong fluorescence at 480 nm. Under a similar condition as described above, we found the formations of thiol adduct with both ThioGlo1 and CPM is linear with respect to the GSH concentration. As shown in **Figure 16**, ThioGlo1 is more sensitive to free thiol with a larger slope. Therefore, we chose ThioGlo1 for the following studies.

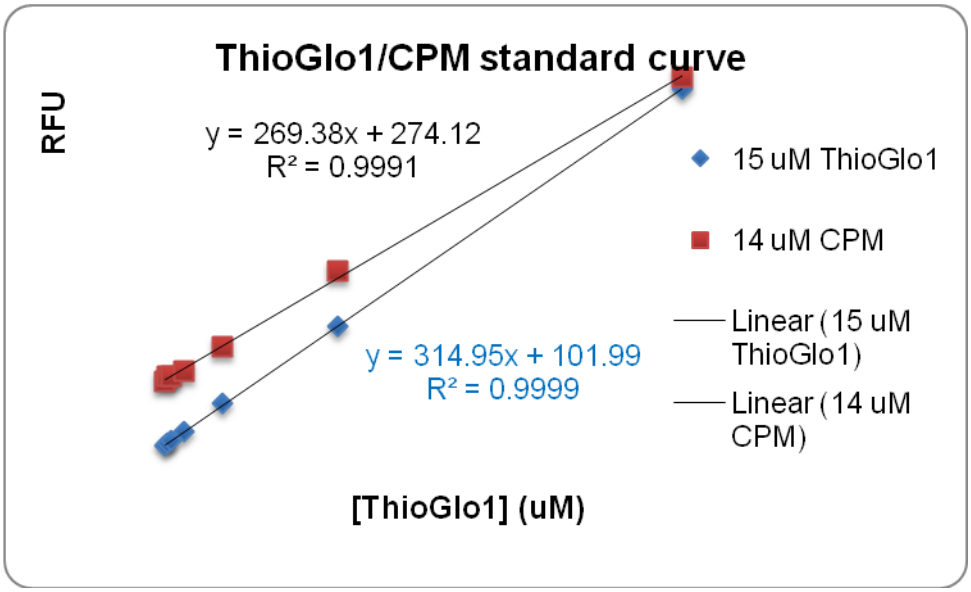


Figure 16. ThioGlo1 *vs*. CPM.

Time-dependent studies

For this study, NTMT1 (0.2 µM) was incubated in the reaction buffer containing 25 mM Tris, 50 mM KCl, pH 7.5, 15 µM ThioGlo1, 10 µM SAHH, 100 µM SAM, and 100 µM RCC1-12 at 37 °C. The concentration of formed product SAH during the reaction was derived from a standard calibration curve generated with glutathione and ThioGlo1. The result of this study indicated a linear relationship between product formation and reaction time during 12 minutes (R^2 = 0.9919) (**Figure 17**). 10 µM SAHH was used to ensure that the hydrolysis of SAH was not rate-limiting as compared to NTMT1 catalysis.

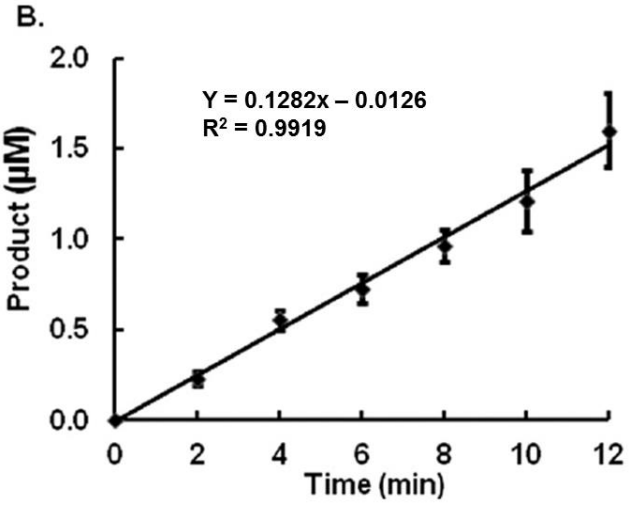


Figure 17. Time dependent studies.[41]

Concentration dependent studies

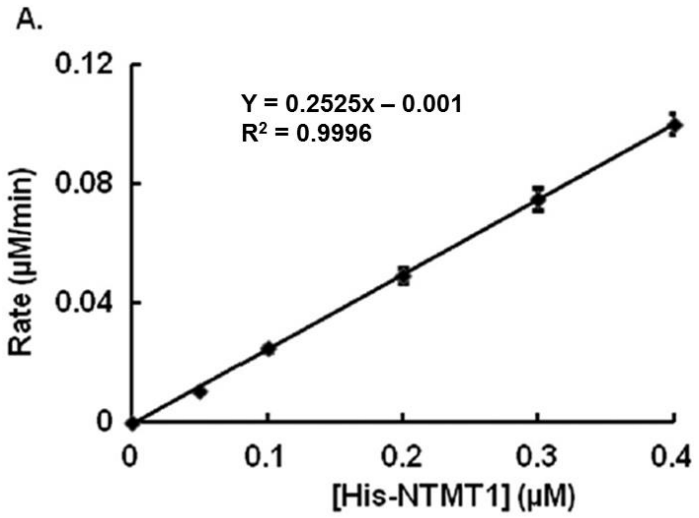


Figure 18. Concentration dependent studies.[41]

In order to determine enzyme concentration and activity linearity range, concentration-dependent studies were conducted in the reaction buffer containing 25 mM Tris buffer, 50 mM KCl, pH 7.5, 15 μM ThioGlo1, 10 μM SAHH, 100 μM SAM, and 100 μM RCC1-12 at 37 °C. Initial velocity was analyzed using a time frame within 10% turnover using Microsoft Excel. Our result showed that the NTMT1 concentration is in linear relationships with reaction rates ranging from 0 to 0.4 μM (R^2 = 0.9996. This result suggests that any enzyme concentration within this range should follow an enzyme concentration-activity linearity relationship. It was decided to use 0.2 μM as the assay concentration for NTMT1 kinetic studies, since it gave a better signal to background ratio (**Figure 18**).

K_m determination of SAM and RCC1-12

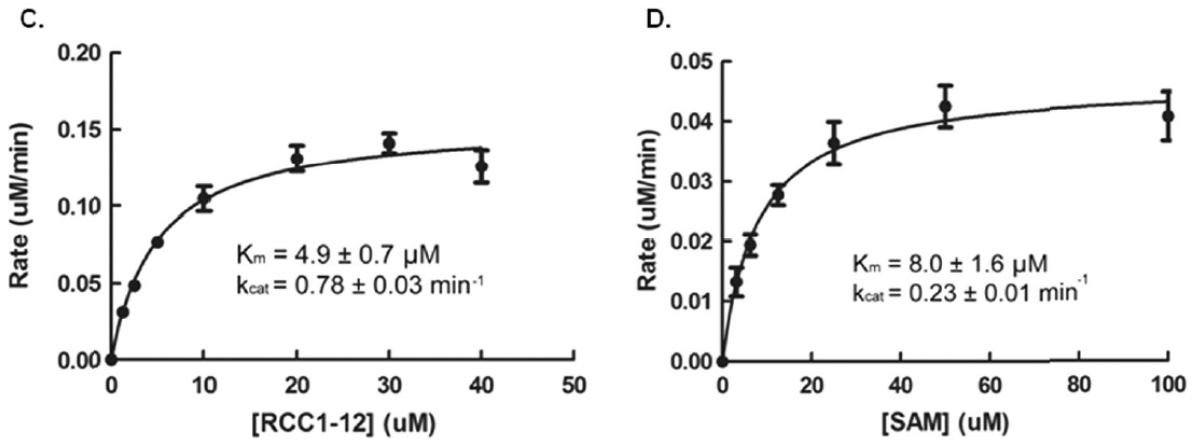


Figure 19. K_m Studies of SAM and peptide substrate RCC1-12.[41]

The steady state kinetic parameters were determined for both RCC1-12 and SAM using our continuous fluorescence assay. As K_m values of SAM for most protein methyltransferases are around 10 μM, we used 100 μM SAM that was assumed to be at a saturated concentration to determine the K_m value of RCC1-12. Various concentrations (0-40 μM) of RCC1-12 peptide were incubated with the reaction mixture and fluorescence was monitored for 12 min. The K_m of RCC1-12 was determined as 4.9 ± 0.7 μM. Likewise, we used 50 μM of RCC1-12 peptide in the presence of various concentration of SAM (0-100 μM) to determine the K_m of SAM, which was 8.0 ± 1.6 μM (**Figure 19**).

2.2 Bisubstrate kinetic mechanism characterization

2.2.1 Design

The data for the initial rates of RCC1-10 peptide were determined at different fixed concentrations of the SAM (12.5, 25, 50, 100 μM). For reactions where the SAM was the varied substrate, the initial rates were examined at fixed concentration of RCC1-10 peptide (1, 2, 4, and 8 μM). The initial rates were globally fit to the following equations using least squares nonlinear regression with GraphPad Prism 5 software.[41]

$$v = \frac{V_{max}[A][B]}{\alpha K^{A} K^{B} + \alpha K^{B}[A] + \alpha K^{A}[B] + [A][B]}$$

When A is saturating, $\alpha K^{B} = K_m^{B}$

$$v = \frac{V_{max}[A][B]}{K^{A} K_m^{B} + K_m^{B}[A] + K_m^{A}[B] + [A][B]}$$

K^{A} and K^{B} are the dissociation constants of the substrate A and B bind to the free enzyme, respectively. V_{max} and K_m are the Michaelis constants.[41]

2.2.2 Lineweaver-Burk double reciprocal plots

Table 2. Initial velocity studies.[41]

varied substrate	fixed substrate	k_{cat} (min^{-1})	$K_{(RCC1)}$ (μM)	$K_{m(RCC1)}$ (μM)	$K_{(SAM)}$ (μM)	$K_{m(SAM)}$ (μM)
RCC1-10	SAM[a]	0.59±0.03	3.3±1.3	1.6±0.4		13.6±2.8
SAM	RCC1-10[b]	1.36±0.09		2.3±0.4	7.85±2.93	6.1±1.7

[a] [SAM] = 12.5, 25, 50, or 100 μM. [b] [RCC1-10] = 1, 2, 4, or 8 μM.

$K_{m(RCC1\text{-}10)}$, $K_{m(SAM)}$, $K_{(RCC1\text{-}10)}$ values were obtained when the concentration of RCC1-10 peptide was varied at different fixed concentrations of the SAM. Similarly, $K_{m(RCC1\text{-}10)}$, $K_{m(SAM)}$, and $K_{(SAM)}$ values were obtained when the concentration of SAM was varied at different fixed concentrations of the RCC1-10 (**Table 2**). The resulting double reciprocal plots exhibit an increasing slope with decreasing SAM and RCC1-10 concentrations respectively, producing intersecting lines with the intercept lying in the second quadrant (**Figure 20**).[41] This pattern indicates that NTMT1 catalysis proceeds in a sequential bi-bi mechanism. Therefore, NTMT1 requires formation of a ternary complex for the initiation of catalytic reaction, which further confirms our rationale of design for bisubstrate inhibitors.

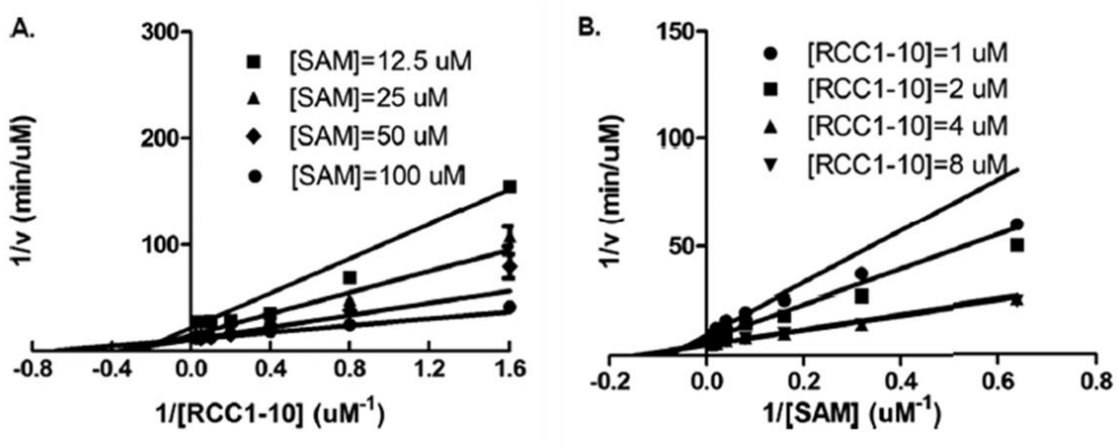


Figure 20. Lineweaver-Burk plot of 1/v *vs.* 1/[RCC1-10] and 1/v *vs.* 1/[SAM].[41]

2.3 NTMT1 methylation progression studies

2.3.1 Design

As discussed previously, substrates of NTMT1 have different methylation states (mono-, di-, and tri-), but it was unknown how di-, or tri-methylation were achieved. If substrates are methylated from the unmethylated state to tri-methylated state in a single step, the methylation progression is following a “processive” mechanism. If substrates were methylated stepwise and intermediates were released during the enzyme catalysis, it is following a “distributive” fashion (**Figure 21)**.

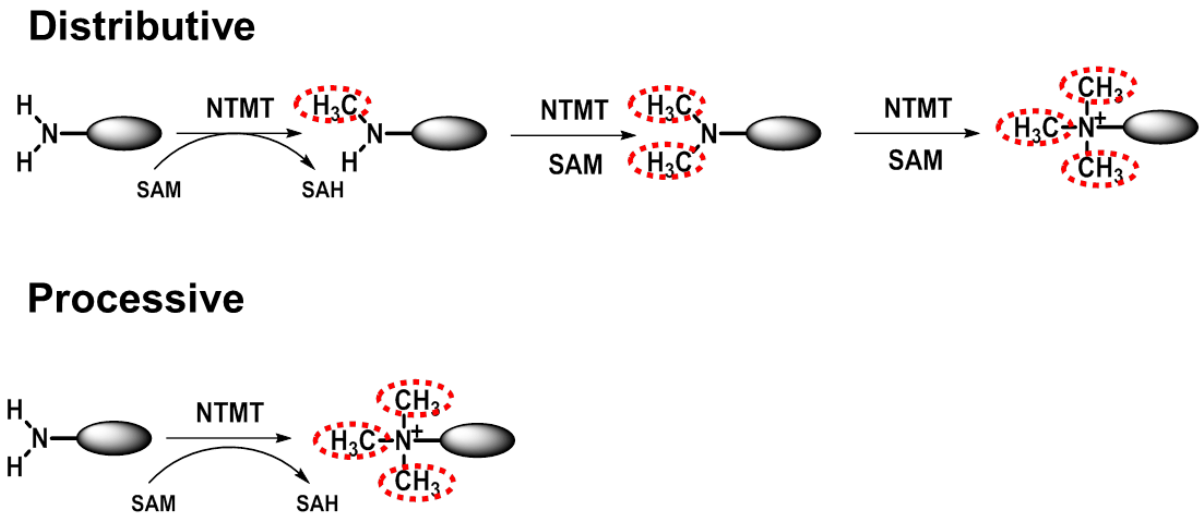


Figure 21. Methylation progression patterns of NTMT1.[41]

2.3.2 Methylation progression studies via MALDI-MS

In order to unveil the methylation progression pattern of NTMT1, Dr. Stacie Richardson from our lab has developed a direct ratiometric quantification, matrix-assisted laser desorption/ionization (MALDI)-MS assay to directly measure substrate concentrations of the varied methylation states.[41] Following this method, Dr. Richardson monitored concentration of different intermediates at different time points (**Figure 22)**. The total concentration of RCC1 used in this experiment was 10 μM. By utilizing the concentration of the RCC1 as an internal standard, and comparing the relative monoisotopic peak areas, we were able to measure the populations of all methylation states simultaneously.

Methylation progression profiles indicate that both Me-RCC1 and Me2-RCC1 had reached 50% of the total substrate population.[41] Thus, 5~6 μM of Me-RCC1 and Me2-RCC1 were detected at different time points during the methylation progression. Since only 0.2 μM of NTMT1 was used in this assay, 5~6 μM of intermediates suggested that intermediates released from NTMT1 are accumulating in the reaction mixture, and subsequently rebound to NTMT1 for further methylation. Overall, the result of methylation progression assay suggests that NTMT1 follows a distributive methylation mechanism. However, a full-length protein substrate may exhibit a processive mechanism if it has significantly higher affinity to the enzyme or a slower off-rate than the enzyme turn-over time. Future study with full-length protein substrates would be interesting to explore this possibility.[41]

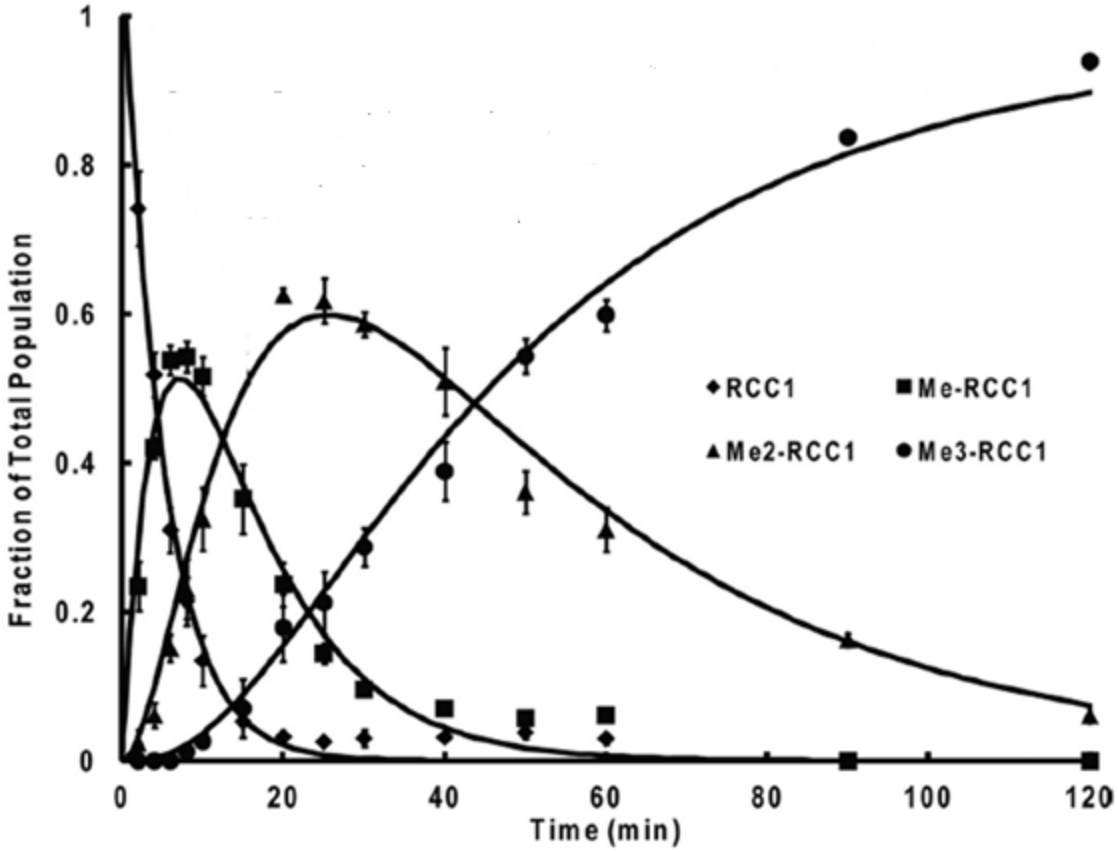


Figure 22. Methylation progression profiles of NTMT1.[41]

2.4 Effects of peptides' length, methylation states, and sequences on substrate binding and recognition

2.4.1 Design

NTMT1 is known to methylate proteins that contain an X-Pro-Lys motif.[70] So far, peptides substrates used for NTMT1 kinetic studies were derived from the first twelve or ten residues of N-terminus of RCC1 since these were reported to have a binding affinity of 10 μM for NTMT1.[79] However, contributions of length, methylation states, and key residues of NTMT1 substrates are still elusive. In order to understand the effects of peptide length regarding enzyme kinetics, we synthesized peptides of varied length (RCC1-6, RCC1-9, RCC1-10 and RCC1-12).

On the other hand, our methylation progression studies suggested that the mechanism of NTMT1 methylation is distributive.[41] It is worth noting that upon each step of methylation, the N-terminus of substrate is sterically changed. So, we prepared peptides with varied methylation states to explore how methylation would affect substrate binding and recognition. To address this question, peptides of varied methylation state were synthesized: RCC1-10, MeRCC1-10, and Me_2RCC1-10. Kinetic studies were carried out on these to determine the kinetic parameters of each peptide.

In order to investigate the contribution and define the specificity of the first residue of the recognition motif (X-Pro-Lys), we chose positively charged Arg, negatively charged Asp, polar aromatic Tyr, and the nonpolar hydrophobic residues Trp and Lys at the first position. Steady state kinetic studies were carried out to determine

the K_m, k_{cat}, and k_{cat}/K_m values of each peptide. ITC analysis was utilized by our collaborator (Dr. Min's lab) to directly analyze their binding affinities.

The second residue Pro at the N-terminus is highly conserved. To understand contributions of Pro regarding substrate binding and recognition, Dr. Min's lab has designed peptides of varied second residues including I, Q, E, and S.

2.4.2 Peptide length effect

Table 3. Kinetic studies of peptide substrates of varied length.[41]

Peptide ID	Sequence	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/Km (M^{-1} min^{-1})
RCC1-6	SPKRIA	3.2 ± 0.4	0.56 ± 0.02	1.8×10^5
RCC1-9	SPKRIAKRR	1.4 ± 0.1	0.53 ± 0.01	3.8×10^5
RCC1-10	SPKRIAKRRS	0.89 ± 0.09	0.44 ± 0.01	4.9×10^5
RCC1-12	SPKRIAKRRSPP	3.1 ± 0.3	0.57 ± 0.01	1.8×10^5

The results of peptides' length studies are summarized in **Table 3**. Kinetic studies on peptides with different lengths (6-12 mer) gave K_m values from 0.81 to 4.72 μM and k_{cat} values from 0.43 to 0.61 min^{-1}. The k_{cat}/K_m values ranged from 1.78×10^5 to 4.93×10^5 M^{-1} min^{-1}. Among these four peptides, RCC1-10 has the lowest K_m of 0.89 ± 0.09 μM, which suggests that a C-terminal residue to the Ser-Pro-Lys motif also contributes to substrate recognition. In comparison, all peptides of different lengths exhibited similar k_{cat}s, which suggests NTMT1 can efficiently catalyze a hexapeptide.

2.4.3 Methylation state effect

Table 4. Kinetic studies of peptide substrates of varied methylation state.[41]

Peptide ID	Sequence	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/Km (M^{-1} min^{-1})
RCC1-10	SPKRIAKRRS	0.89 ± 0.09	0.44 ± 0.01	4.9×10^5
MeRCC1-10	Me-SPKRIAKRRS	1.4 ± 0.1	0.58 ± 0.01	4.1×10^5
Me2RCC1-10	Me2-SPKRIAKRRS	4.3 ± 0.5	0.59 ± 0.02	1.4×10^5

Since RCC1-10 exhibited the lowest K_m (**Table 3**), this peptide was chosen for the studies of methylation effects (**Table 4**). RCC1-10, MeRCC1-10, and Me_2RCC1-10 represent three different methylation states. The kinetic studies indicated that unmethylated (RCC1-10) and mono-methylated (MeRCC1-10) peptides have comparable K_ms of 0.89 μM and 1.4 μM, respectively. In contrast, di-methylated peptide has a four-fold increased K_m compared to the unmethylated peptide, which suggests that NTMT1 can bind and catalyze unmethylated and mono-methylated RCC1-10 in a similar fashion. But the di-methylated peptide may have introduced steric factors to an extent that affects substrate binding.

While comparing k_{cat} of each peptides, it suggests that di-methylated RCC1-10 has comparable turnover numbers compared to unmethylated and mono-methylated peptide, which further proved our hypothesis that varied methylation state can significantly affect substrate binding, yet NTMT1 can still efficiently catalyse peptides of different methylation state.

2.4.4 Effects of the first residue

As the hexapeptide was shown to be efficiently methylated by NTMT1 in our previous study, our collaborator (Dr. Min's lab) prepared a series of hexapeptides to investigate the contribution and the tolerability of the first residue of the N-terminus: SPKRIA (hRCC1), PPKRIA (mRCC1), RPKRIA (positively charged), YPKRIA (polar aromatic), WPKRIA (non-polar hydrophobic), LPKRIA (non-polar hydrophobic), and DPKRIA (negatively charged).

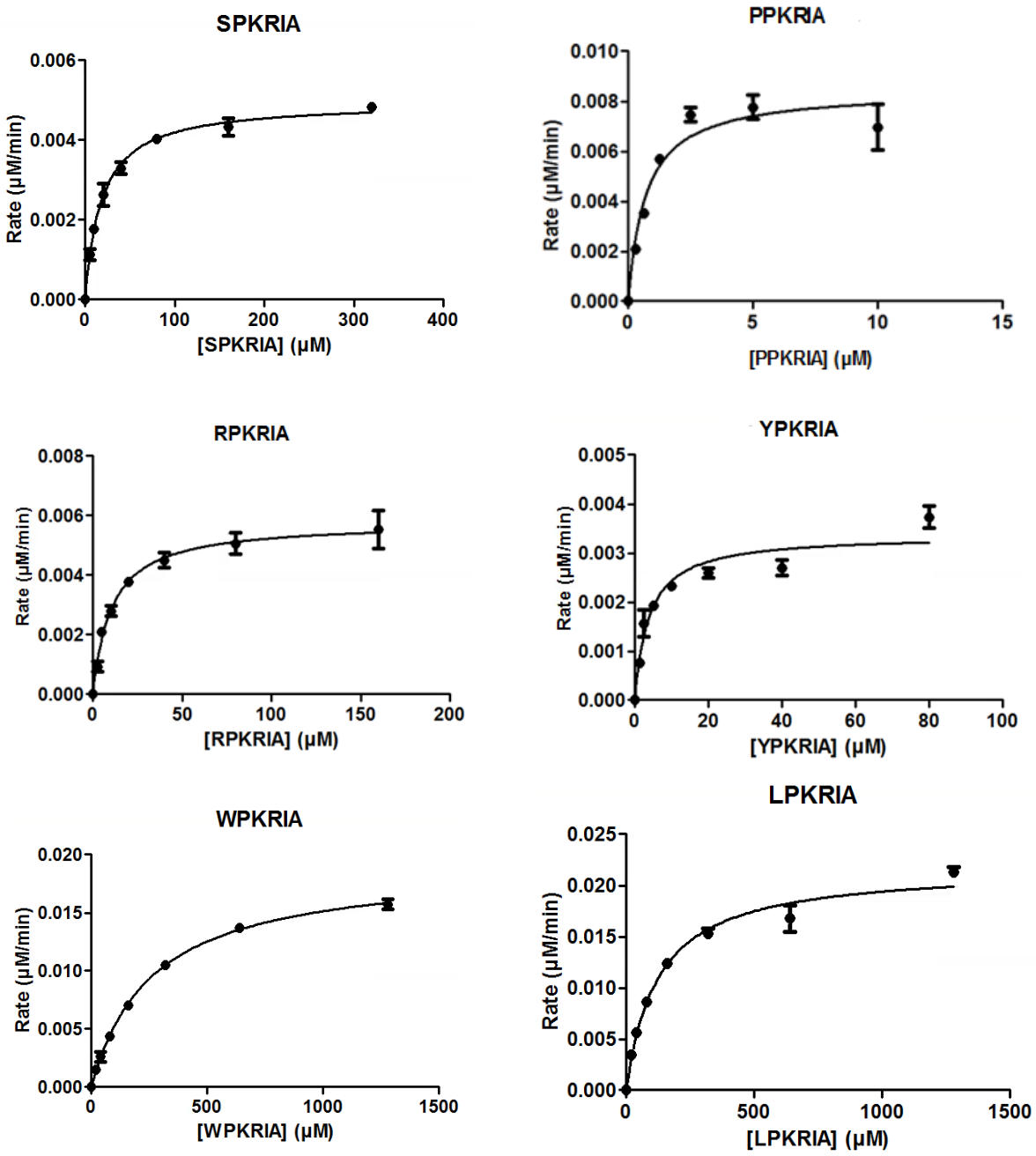


Figure 23. Steady state kinetic profiles of peptides with varied first residue.[102]

Table 5. Steady state kinetic studies of peptides with varied first residue.[102]

Peptide	Sequence	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}min^{-1}$)
RCC1-6-1S	SPKRIA	7.9 ± 0.7	0.07 ± 0.001	9.0×10^{-3}
RCC1-6-1P	PPKRIA	0.3 ± 0.04	0.11 ± 0.006	3.8×10^{-1}
RCC1-6-1Y	YPKRIA	1.6 ± 0.3	0.04 ± 0.003	2.5×10^{-2}
RCC1-6-1R	RPKRIA	4.0 ± 0.5	0.09 ± 0.003	2.3×10^{-2}
RCC1-6-1W	WPKRIA	126 ± 7	0.10 ± 0.002	7.9×10^{-4}
RCC1-6-1L	LPKRIA	54 ± 6	0.11 ± 0.004	2.0×10^{-3}
RCC1-6-1D	DPKRIA	ND	ND	ND

(ND) No detectable activity at 250 μM peptide and a saturating amount of SAM.

Steady state kinetic profiles of each peptides substrate are presented in **Figure 23**. Results of kinetic studies of these peptides with different first residues are summarized in **Table 5**. As indicated, despite the fact that the only structural variation of these peptides is the sidechain of the first residue, their K_m values span from 0.3 μM (PPKRIA) to 126 μM (WPKRIA). Compared to SPKRIA (K_m = 7.9 ± 0.7 μM), both RPKRIA and YPKRIA have 2 ~ 4 fold lower K_ms of 4.0 ± 0.5 μM and 1.6 ± 0.3 μM, respectively. This is in agreement with previous reported studies that the first residue can be positively charged due to the extensively negatively charged substrate binding site (see Appendix).[79] A positively charged residue like arginine is electronically favorable. In addition, from our crystal structure, there is a spacious pocket by the

sidechain of the first residue, the size of which can accommodate an aromatic ring such as tyrosine.[102]

In comparison, the non-polar hydrophobic peptides WPKRIA and LPKRIA have significantly increased K_ms of 126 $\pm$ 7 μM and 54 $\pm$ 6 μM, respectively, which suggests that the substrate binding site of the first residue is unfavorable with respect to non-polar hydrophobic residues. Lastly, the negatively charged peptide DPKRIA did not show any sign of methylation; this is in agreement with a previous study that negatively charged binding site tend to dispel negatively charged residues.[79]

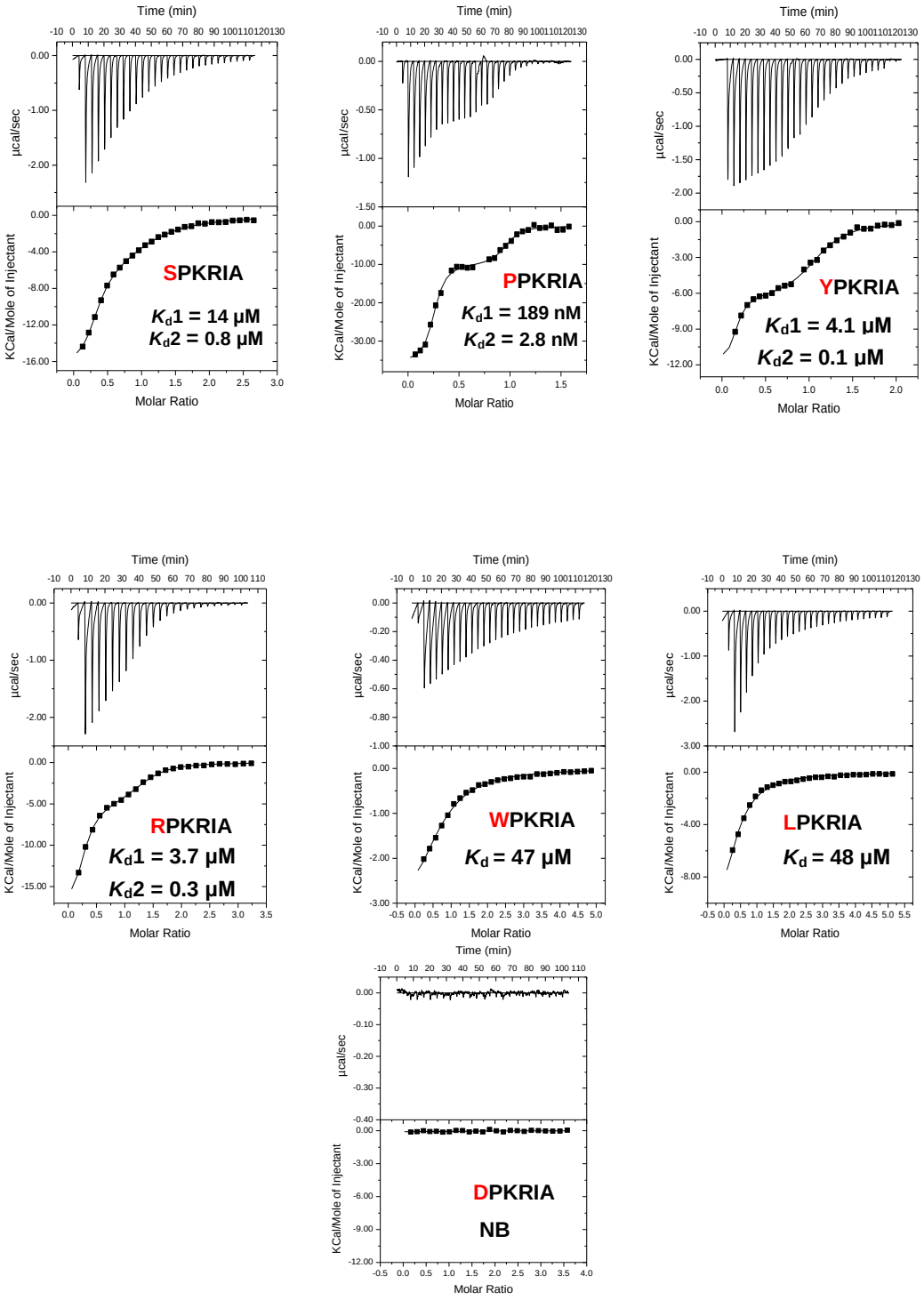


Figure 24. ITC analysis of peptides with varied first residues.[102]

Results of ITC analysis are in agreement with that of the kinetic studies. PPKRIA shows the best binding with K_d = 189 nM. RPKRIA (K_d = 3.7 uM) and YPKRIA (K_d = 4.1 μM) have about three folds lower K_d values compared to SPKRIA (K_d = 14 μM). Both hydrophobic residues WPKRIA (K_d = 47 μM) and LPKRIA (K_d = 48 μM) have 3-fold increased K_d compared to SPKRIA. Negatively charged DPKRIA did not show any sign of binding (**Figure 24**).

Interestingly, a bimodal binding was observed from the ITC profiles of some peptides (**Figure 24**).[102] This is especially evident for those peptides with high binding affinities, such as PPKRIA, RPKTIA, and YPKRIA. We hypothesized that it may be attributed to the possibility that NTMT1 purified from *E. coli* contains endogenous SAM, which could consequently methylate peptides during ITC analysis. Co-purification of endogenous SAM along with methyltransferase has been documented before.[103] Therefore, the observed bimodal binding is resulted from a mixture of methylated and unmethylated substrates with their varied binding affinities for NTMT1. In order to test our hypothesis, the MALDI-MS based assay was carried out without external addition of SAM. As mentioned before, N-terminal methylation of a substrate requires the formation of a ternary complex of NTMT1–peptide–SAM. If NTMT1 used in the experiment does not contain SAM, no methylation should be observed, and vice versa.

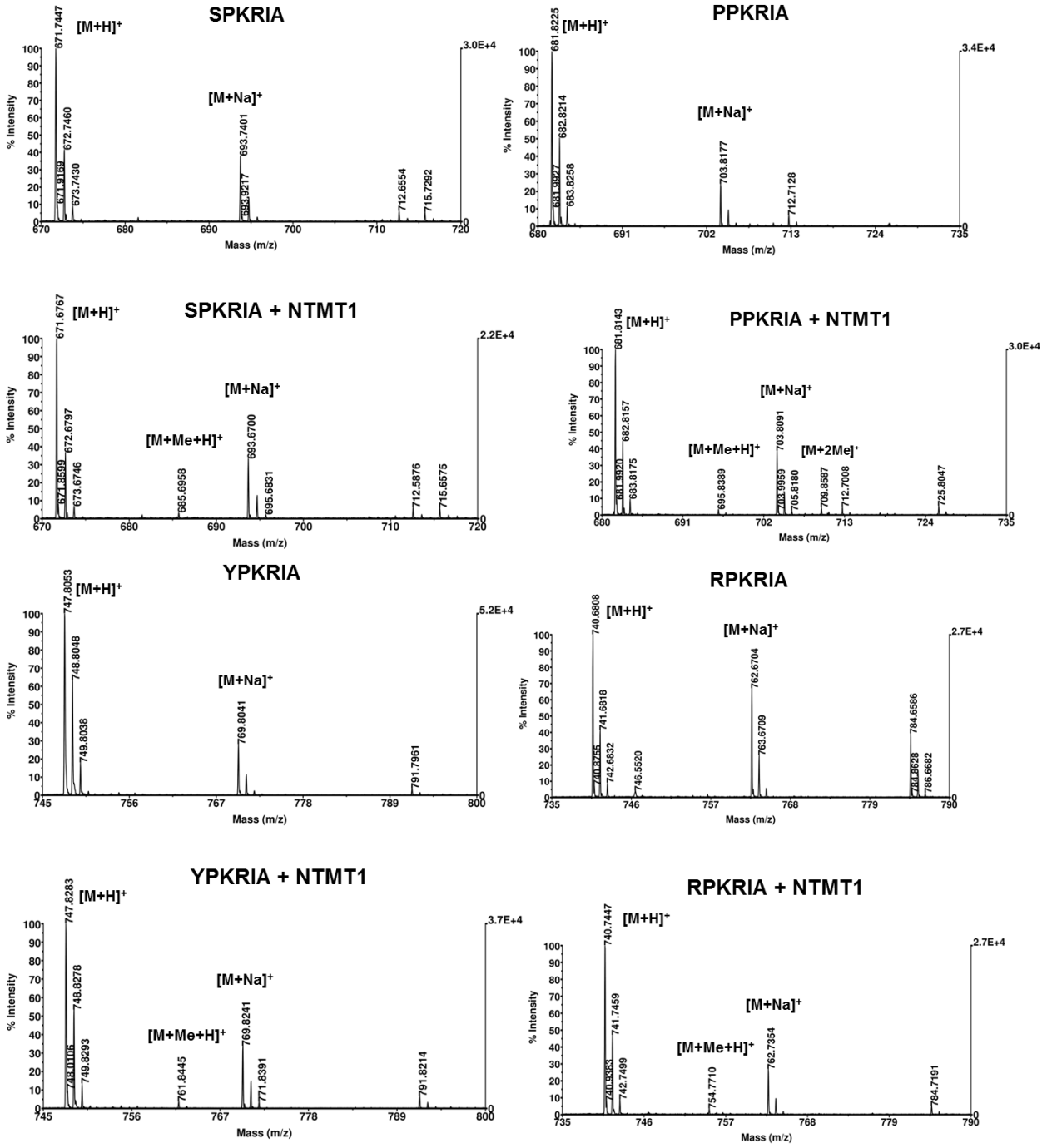


Figure 25. MALDI-MS spectra of S/P/Y/RPKRIA.[102]

Results of MALDI-MS based study are shown in **Figure 25**. Spectra that are named by peptide sequences are controls with peptides alone, from which only $[M + H]^+$ and $[M + Na]^+$ peaks are observed. However, spectra labelled with peptide sequence

plus NTMT1 contain $[M + Me + H]^+$ peaks. Peptide with high binding affinity such as PPKRIA even contains the $[M + 2Me]^+$ peak. The above results from MALDI–MS based assay supports our hypothesis that NTMT1 used in the experiment contains a certain amount of endogenous SAM, which causes the bimodal binding phenomenon observed from ITC analysis studies.[102]

2.4.5 Effects of the second residue

Four peptides of varied second residues were synthesized: SIKRIA, SQKRIA, SEKRIA, SSKRIA. Kinetic studies were carried out to determine their kinetic parameters.

Table 6. Steady state kinetic study results of peptides with varied second residues.[102]

Peptide	Sequence	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}min^{-1}$)
RCC1-6-2I	SIKRIA	ND	ND	ND
RCC1-6-2Q	SQKRIA	ND	ND	ND
RCC1-6-2E	SEKRIA	ND	ND	ND
RCC1-6-2S	SSKRIA	ND	ND	ND

(ND) No detectable activity at 250 μM peptide and a saturating amount of SAM.

As summarized in **Table 6**, after replacement of the second Pro by Ile, Gln, Glu, and Ser, no methylation could be detected from fluorescence assay. Results from ITC analysis are consistent with results from kinetic studies, which are shown in **Figure 26**. Both kinetic studies and ITC analysis suggested that Pro2 is essential as the binding abilities of all four peptides are completely abolished. Previously reported substrate specificity studies of NTMT1 stated that the second residue Pro can be replaced by Ala, Glu, Met, Asn, Gln, Gly and Ser through *in vitro* peptide methylation assays and substrate immunoprecipitations.[79] However, our studies were the first quantitative

measurement of various peptide substrates and results supported the dogma of X-Pro-Lys motif.[102]

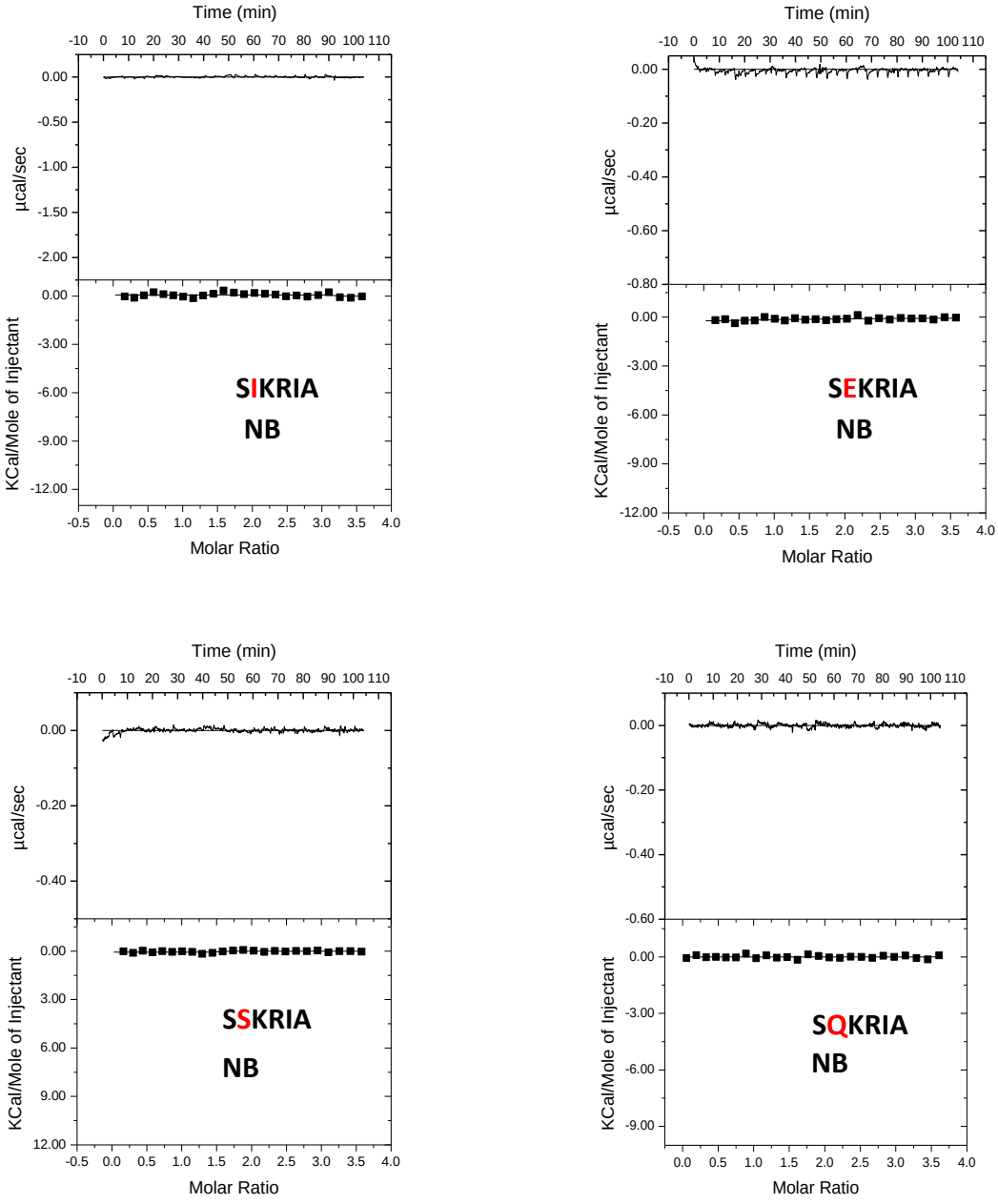


Figure 26. ITC analysis of peptides with varied second residues.[102]

2.5 Structural basis study for substrate binding and recognition

2.5.1 Design

New co-crystal structures of NTMT1 in complex with SAH and two different hexapeptide substrates including the human (hRCC1-6: SPKRIA) and mouse RCC1 (mRCC1-6: PPKRIA) were successfully obtained by **Dr. Jinrong Min**'s lab at the Structural Genomics Consortium.[102] These crystal structures reveal that its substrate peptides are inserted into a negatively charged channel of NTMT1 (see Appendix), which is in striking contrast to that of other protein methyltransferases (see Appendix). We identified a few key residues (N168, W136, D180 and D177) that contribute to the substrate recognition and performed site-directed mutagenesis studies to elucidate the interactions (**Figure 27**).[102]

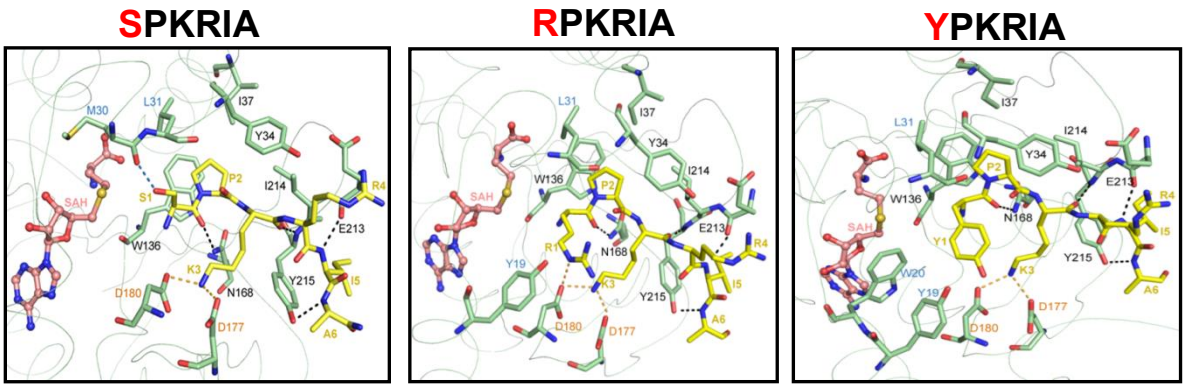


Figure 27. NTMT1–SAH–R/Y/SPKRIA ternary complex.[102]

2.5.2 Kinetic studies of NTMT1 mutants

The carboxamide group of the side chain of Asn168 interacts with the backbone carbonyl group of Ser (**Figure 27**). To examine the contribution of Asn168,

site-directed mutagenesis was carried out to replace Asn168 with a Lys. The results of kinetic studies that characterized through the fluorescence assay (**Table 7)** have demonstrated that N168K has ~36-fold higher of K_m and two-fold lower of k_{cat} than wtNTMT1.[102] Therefore, this mutation has affected not only substrate binding and recognition, but also enzyme catalytic activity.

Trp136 was identified as having a stacking interaction with the sidechain of Pro2. After mutation of Trp136 to Phe, the kinetic parameters dramatically decreased with K_m > 200 μM and k_{cat} > 0.04 $\pm$ 0.01 min^{-1} (**Table 7**). Meanwhile, mutant of Trp136 to Ile has undetectable enzyme activity. Both mutants suggest the importance of the stacking interaction between side chains of Pro2 and Trp136.[102]

Table 7. Steady state kinetic studies of NTMT1 mutants.[102]

Enzyme	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}min^{-1}$)
Wide type	7.3 $\pm$ 0.7	0.10 $\pm$ 0.01	1.4×10^{-2}
N168K	263 $\pm$ 141	0.05 $\pm$ 0.01	1.9×10^{-4}
W136F	>200	>0.04 $\pm$ 0.01	/
W136I	ND	ND	ND
D180K	ND	ND	ND
D180Y	ND	ND	ND

(ND) No detectable activity at 250 μM peptide and a saturating amount of SAM.

To examine the essentiality of the two interactions between Lys3 and Asp180/Asp177 (**Figure 27**), site directed mutagenesis was carried out to mutate both residues. Those mutations were expected to disrupt these electrostatic interactions. Our results from kinetic studies indicate that neither D180K nor D180Y have detectable methylation activity (**Table 7**). While integrating this result with the ITC analysis that

performed by our collaborator, Dr. Min's lab (**Figure 28**), it can be seen that the peptide substrate (SPKRIA) is not able to bind to either D180K or D180A, which further confirmed that the interaction between Asp180 and Lys3 is essential in terms of substrate binding. On the other hand, compared to wtNTMT1 (K_d =14 μM), D177A has about two-fold increased K_d (32 μM), which is an indication of the decreased binding affinity after interrupting the interaction between Asp177 and Lys3.[102]

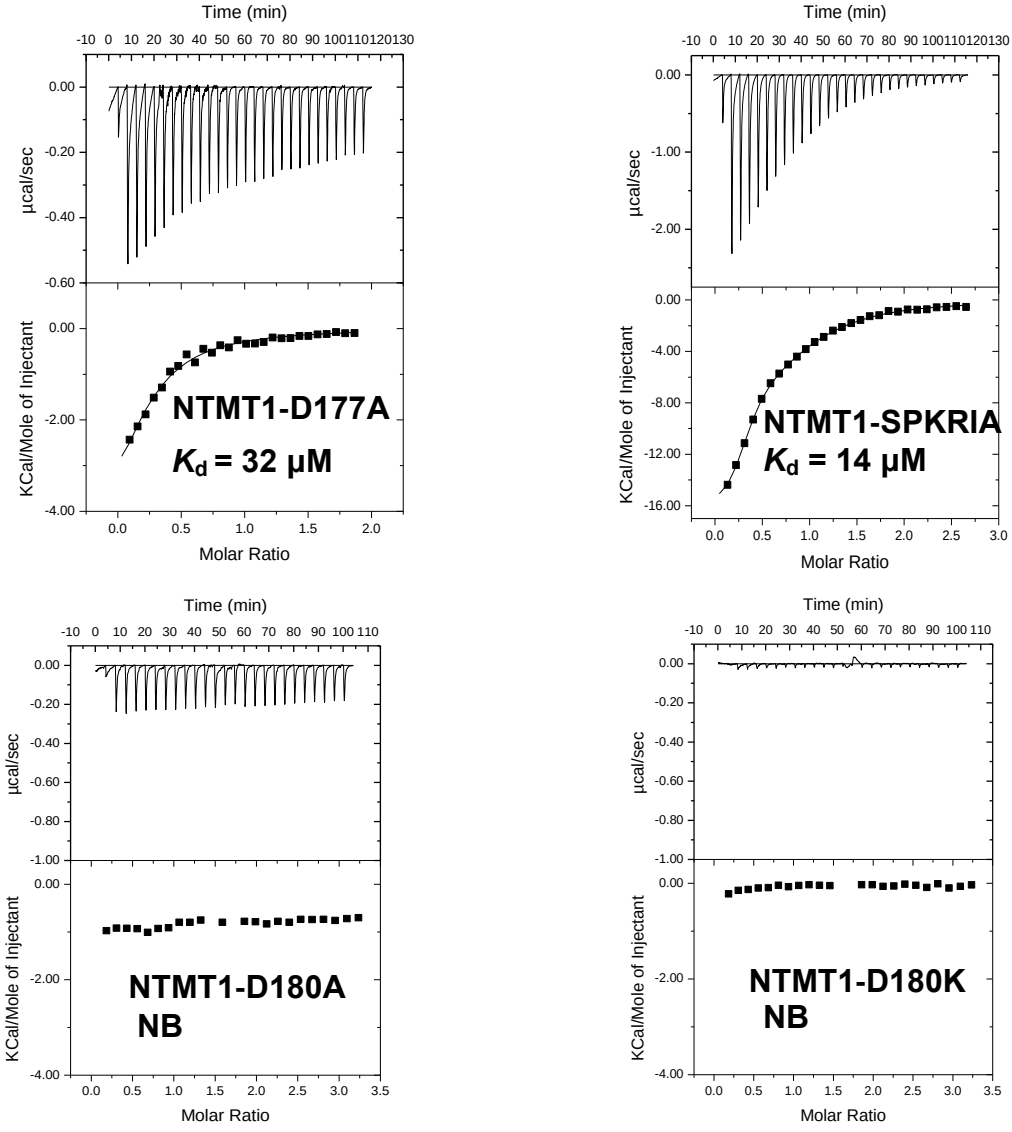


Figure 28. ITC analysis of mutants D180A, D180K and D177A.[102]

2.6 Discovery of a small molecule inhibitor for NTMT1

2.6.1 Design

Specific chemical probes that modulate methylation have functioned as valuable chemical tools to investigate NTMT1-mediated biological processes. Our lab developed the first potent and specific NTMT1 bisubstrate inhibitor that displays an IC_{50} of 0.8 μM for NTMT1 and is more than 60-fold selective over protein lysine methyltransferase G9a and arginine methyltransferase 1.[101] The bisubstrate inhibitors offered us a valuable probe for biochemical studies of NTMT1, but have limited use in biological studies due to their low stability in the presence of serum in cell culture media. While we continue our efforts to optimize structures of our bisubstrate inhibitors, discovery of novel small molecule inhibitors for NTMT1 will be critical for the interrogation of the biological functions of NTMT1 and elucidatation of the different responses to knockdown of NTMT1 in normal and transformed cells.

We used the only available (at that time) X-ray crystal structure for the human NTMT1 protein in complex with product SAH (PDB ID: 2ex4, 1.75 Å) to predict the substrate binding site and develop the initial search query in collaboration with the Dr. Glen Kellogg's lab. Our goal was to discover small molecule inhibitors that target the binding site of the peptide substrate; thus, the active site was defined by the cavity opposite the SAH binding site. The flow chart of this study is illustrated in **Figure 29**.

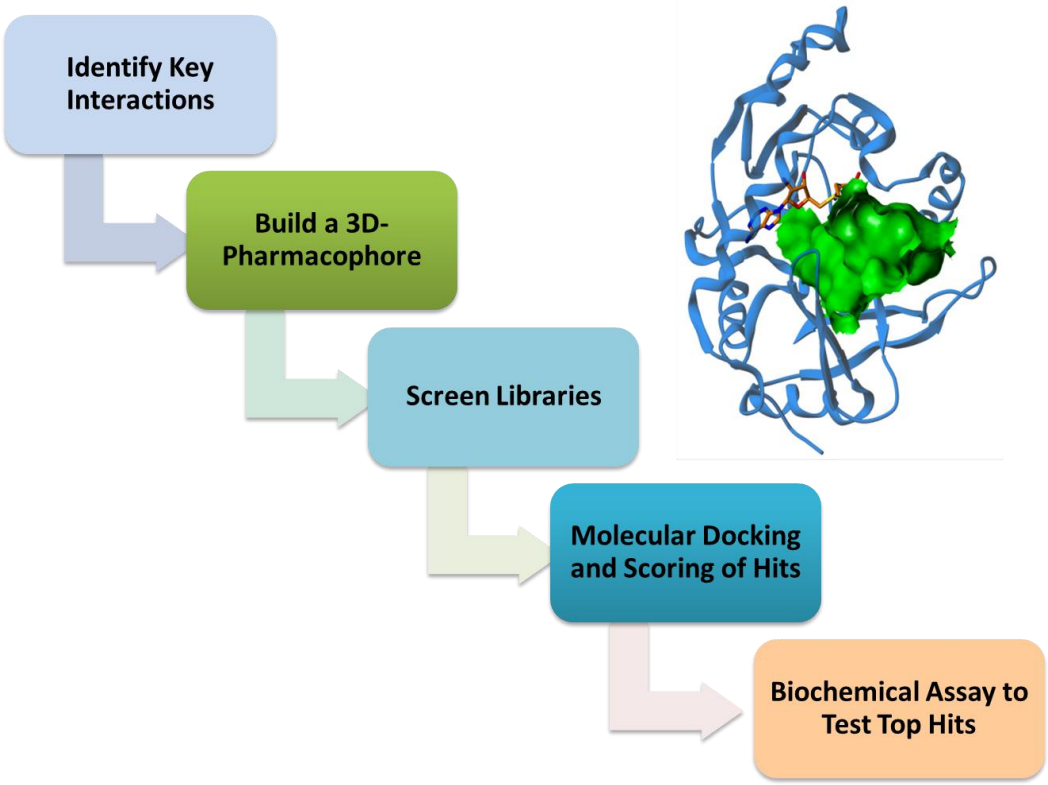


Figure 29. Project plan for discovery of small molecule inhibitors.

2.6.2 Computational studies

We defined the query based on several features surrounding the pocket, including one aromatic site (Trp136), two acceptor sites (Asp177, Asp180), and one donor site (Asn168). The NTMT1 structure, including the SAH, was processed in Sybyl-X 1.1 by deleting all water molecules, adding all protons and performing an energy minimization of these protons while holding the heavy atoms as an aggregate. Initial virtual screening was performed with the UNITY module of Sybyl against the open NCI database of 250,000 unique compounds. The resulting 150 UNITY hits were then docked with GOLD 5.1 and the highest scoring poses were then energy minimized and rescored with HINT. These 150 promising leads were ranked based on HINT score and were visually inspected. We requested from the NCI these high-ranking compounds for primary screening with the fluorescence-based assay.

2.6.3 Biochemical screening with recombinant NTMT1

Single dose inhibitory activity studies

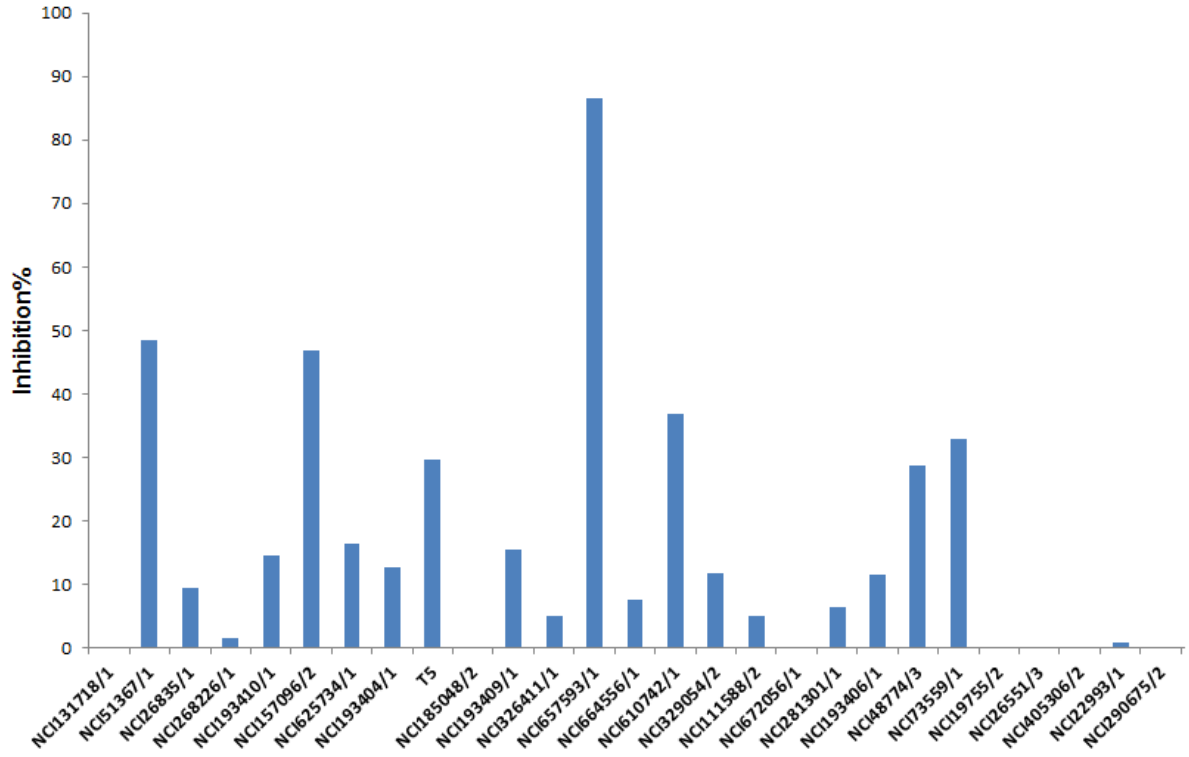


Figure 30. Results of 100 μM primary screening.

We aim to identify selective small molecule inhibitors that target the unique peptide substrate binding site of NTMT1, so we used an excess amount of SAM (100 μM) and substrate peptide RCC1-12 peptide at its K_m value of 5 μM in the screening. The reaction was allowed at 37 ^{o}C for 10 min with or without the addition of inhibitors. Obtained compounds were subject to single dose screening. 100 μM was used as the final inhibitor concentration. Results of single dose screening are shown in **Figure 30**. Compound NCI657593 was identified exhibiting over 85% inhibition in 100 μM. The

resulting top five compounds were chosen (NCI657593, NCI51367, NCI610742, NCI48774, and NCI73559) for subsequent studies.

Secondary assays to remove false positives

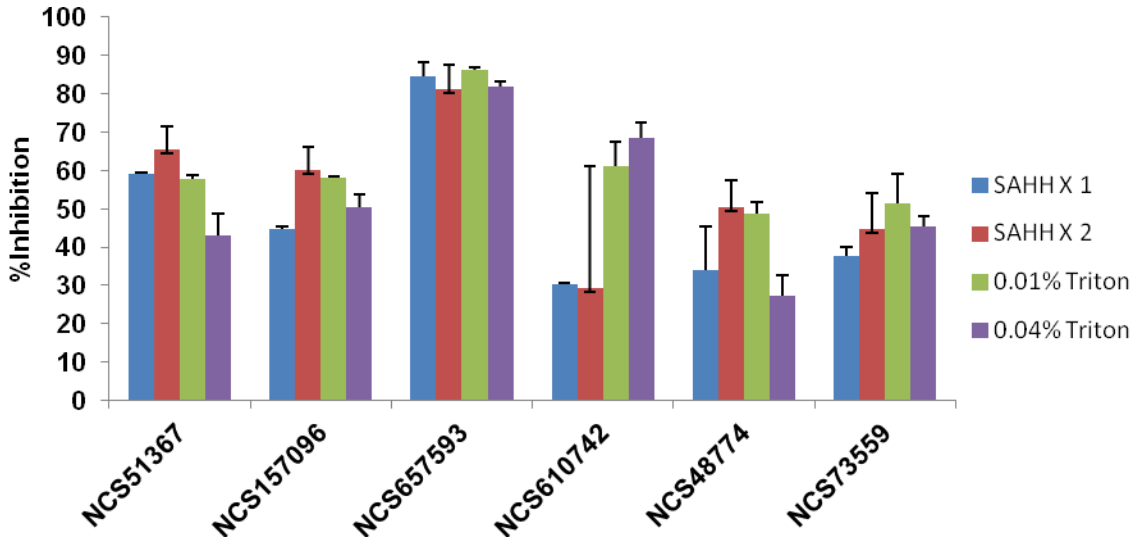


Figure 31. Secondary screening to remove false positive.

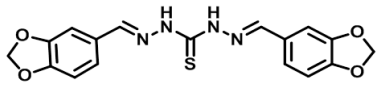
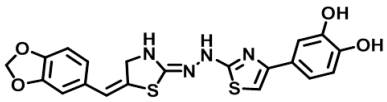
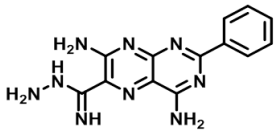
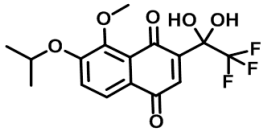
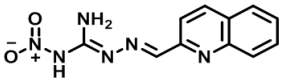
This experiment was conducted in order to exclude possibilities of false results. If an inhibitor is targeting SAHH, increasing SAHH concentration should show reduced inhibitory activity. Results from **Figure 31** indicate that after doubled SAHH concentration, none of our compounds showed drastic changes in terms of inhibitory activity, and this suggests that their activity is not from SAHH inhibition.

Occasionally, introduction of a compound to the assay can result in protein aggregation. Aggregated enzymes normally have no enzymatic activity, which also leads to false inhibition results. Triton is a detergent commonly used in biochemical

assay to prevent protein aggregation. Through addition of triton, if the inhibitory activity of a compound is caused by aggregation, it would show reduced inhibitory activity. Our results indicate that NCI48774 showed significant reduced inhibitory activity after 0.04% of triton was added, which implies that this compound might introduce protein aggregation, and its inhibitory result is not reliable.

IC_{50} studies

Table 8. IC_{50} studies.

	51367 (IC_{50}: 115.3 ± 34.4 μM)
	657593 (IC_{50}: 7.4 ± 0.6 μM)
	73559 (IC_{50}: 159.7 ± 14.9 μM)
	610742 (IC_{50}: 104.3 ± 3.9 μM)
	157096 (IC_{50}: 73.4 ± 4.8 μM)

IC_{50} is an important indicator of the potency of an inhibitor. In this study a three-time serial dilution of each sample was carried out with DMSO. The inhibitory activities of compounds in different concentrations were monitored through fluorescence assay. The results of IC_{50} study indicate NCI657593 has the lowest IC_{50} of 7.4 ± 0.6 μM. Other compounds have at least >70 μM IC_{50} (**Table 8**).

Selectivity study

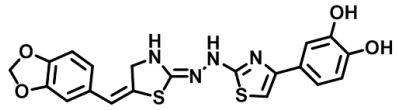
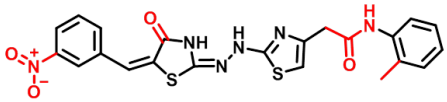
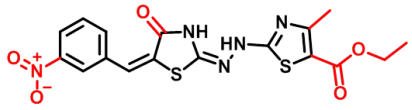
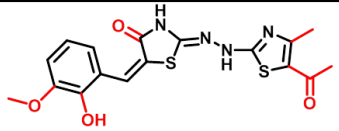
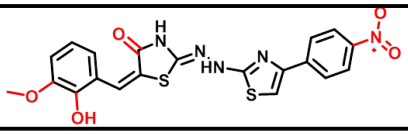
Table 9. Selectivity study.

	IC_{50} (μM)		
Compound	**NTMT1**	**PRMT1**	**G9a**
NCI657593	**7.4 ± 0.6 μM**	**>50 μM**	**>50 μM**

Inhibitor selectivity was determined by comparing IC_{50} values of the inhibitor with other methyltransferases; e.g. PRMT1 and G9a. As shown in **Table 9**, compared to NTMT1, compound NCI657593 has >50 μM IC_{50} for both arginine methyltransferases PRMT1 and lysine methyltransferase G9a, which suggests that it is selective for NTMT1 among these three methyltransferases.

Similarity search

Table 10. Inhibitory activity and IC_{50} studies of NCI657593 analogues.

		IC_{50}	Inhibition% (100uM)
NCI657593		7.4 ± 0.6 uM	83.2 ± 2.5%
NCI657586		90.6 ± 6.3 uM	42.6 ± 0.6%
NCI657569		70.2 ± 4.8 uM	53.4 ± 1.0%
NCI659424		86.5 ± 11.4 uM	62.9 ± 2.2%
NCI659531		111.14 ± 13.4 uM	48.8 ± 1.5%

Since NCI657593 showed extraordinary inhibitory activity, compounds that are structurally similar to NCI657593 were ordered. Subsequently, 100 μM single dose inhibition study and IC_{50} studies were conducted to examine their inhibitory activity. However, results indicate that all compounds possess >70 μM IC_{50} (**Table 10**). Even though NCI657593 was identified as a small molecule with significant inhibitory activity and selectivity, the molecular structure of NCI657593 remains to be validated through synthesis or further characterization.

2.7 Substrate specificity and methylation progression studies of NTMT1/NTMT2

2.7.1 Project design

Along with the discovery of NTMT1 (METTL11A), a homologue named METTL11B (NTMT2) was also identified.[81] NTMT2 recognizes the same N-terminal motif of X-P-K/R, and possesses N-terminal methylation activity.[85] Structurally, NTMT2 has about 50% sequence identity and 75% sequence similarity with NTMT1 (**Figure 32**). Those residues that were identified crucial for substrate binding and recognition in the substrate binding site (Trp136, Asn168, Asp177, Asp180) are also conserved in NTMT2.[85] In addition, it has similar cellular localization and tissue expression pattern as NTMT1. However, as described previously, NTMT2 was reported as a monomethylase to prime the substrates of NTMT1 to facilitate further methylation of substrates by NTMT1.

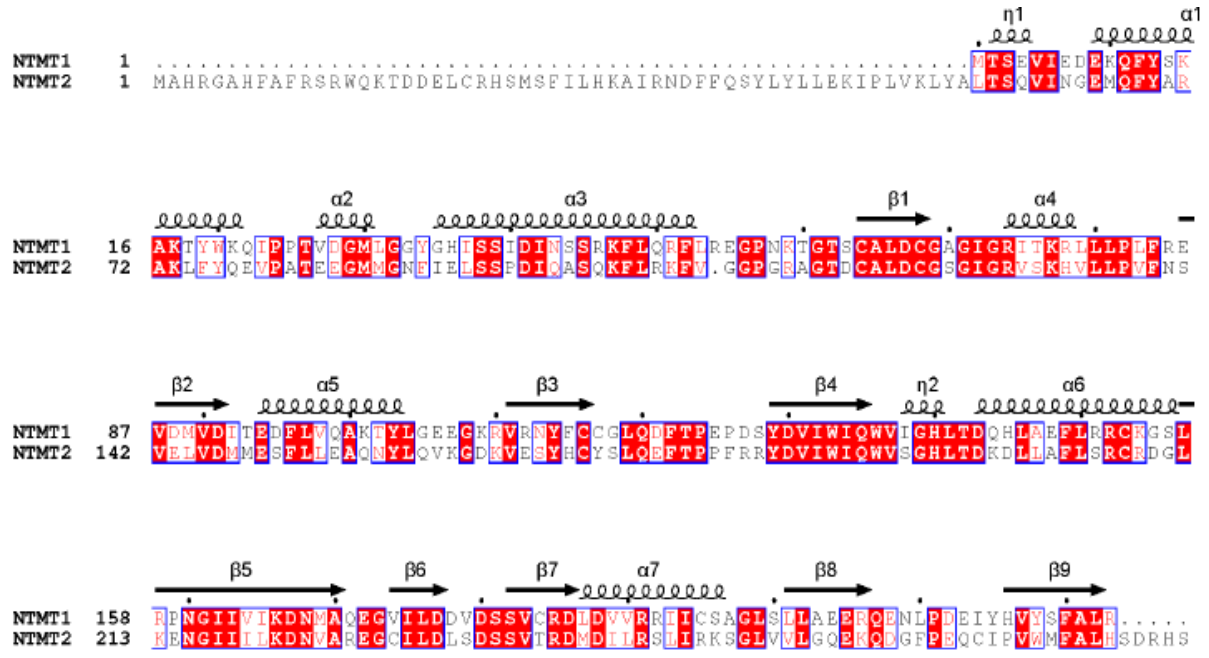


Figure 32. Sequence alignment of NTMT1 and NTMT2.

The goal of this study is to analyze substrate and product specificity of NTMT1 and NTMT2 to understand the role and function of NTMT2 and its relationship with NTMT1. Due to the low activity of NTMT2, we were not able to characterize the kinetic parameters through fluorescence assay. Hence we used a MALDI-MS based assay to monitor methylation progression for substrates of varied first residues with NTMT1 and NTMT2 in parallel.

2.7.2 Methylation progression studies

Table 11. Methylation progression study results summary.

Peptide	NTMT1/NTMT2		
	Mono-methylation	**Di-methylation**	**Tri-methylation**
SPKRIA	YES/YES	YES/NO	NO/NO
Sme1PKRIA		YES/NO	YES/NO
PPKRIA	YES/YES	YES/NO	
Pme1PKRIA			YES/YES
YPKRIA	YES/YES	YES/NO	NO/NO
RPKRIA	YES/YES	YES/NO	NO/NO
WPKRIA	YES/YES	YES/NO	NO/NO
DPKRIA	NO/NO	NO/NO	NO/NO
QPKRIA	YES/YES	YES/NO	NO/NO
NPKRIA	YES/YES	YES/NO	NO/NO
LPKRIA	YES/YES	YES/NO	NO/NO
GPRRRS	YES/YES	YES/YES	YES/YES
GPKRRQ	YES/YES	YES/YES	YES/YES

The results of these methylation progression studies of NTMT1 and NTMT2 are summarized in **Table 11**. The remaining mass spectra of the other peptides are attached in Appendix (Figure 1 ~ Figure 26). As shown, both NTMT1 and NTMT2 can

monomethylate most peptides except DPKRIA. From our previous studies, DPKRIA shows no binding to NTMT1, probably because the negatively charged aspartic acid is considered unfavorable for similarly negatively charged substrate binding site.

The difference between both enzymes starts to show at the di-methylation step. It indicates that, except peptides GPKRRQ and GPRRRS (and DPKRIA), all peptides can be di-methylated by NTMT1. However, none of them achieved di-methylation through NTMT2. It is also worth noting that from time-dependent methylation progression studies, NTMT2 showed an apparent slower methylation rate than NTMT1. Therefore, the di-methylation incapability of NTMT2 is whether due to its slower methylation rate, or another intrinsic mechanism remains to be discovered.

On the other hand, our results indicate that peptides with high binding affinity such as GPKRRQ, GPRRRS, and MePPKRIA had illustrated full methylation (tri-methylation of GPKRRQ, GPRRRS, and di-methylation of Me-PPKRIA) by both enzymes.

3. Conclusions

We have characterized the kinetic mechanism of recombinant NTMT1 using a fluorescence assay and mass spectrometry. The results of studying the initial velocity indicate that methylation by NTMT1 proceeds via a random sequential bi-bi mechanism, which implicates that both SAM and peptide substrate need to bind to NTMT1 to form a ternary complex to initiate enzymatic reaction. This mechanism supports our rationales of designing bisubstrate inhibitors that mimic the ternary complex during enzyme catalysis. Thus compound could simultaneously inhibit both binding sites. In addition, our processivity studies demonstrate that NTMT1 proceeds via a distributive mechanism for multiple methylations. Our processivity studies had indicated the existence of mono- and di-methylated substrates of significant concentrations while enzyme catalysis reactions, which indirectly suggests that both mono- and di-methylated substrate play roles in the biological system.

The results of peptides' length, methylation states, and sequence on substrate recognition showed that hexapeptides gave comparable K_m and k_{cat} values. The K_m studies for peptides with varied N-terminal residues defined that NTMT1 can recognize a motif X-P-K/R, where X can be any amino acid except D/E. Newly identified substrates like centromere H3 variants (CENP-A/B) supported this expanded consensus. We also determined that conserved residues N168, W136, D117, and D188 are critical for substrate binding. These results along with the results from substrate specificity studies

had provided valuable information in terms of substrate binding and recognition in structural basis, which lay a solid foundation for designing of highly selective peptidomimetic and/or small molecule inhibitors in the future.

In the meantime, we utilized computational studies and fluorescence assays to discover a small molecule inhibitor that targets the substrate binding site of NTMT1. We identified one compound that exhibits potent inhibitory activity with an IC_{50} value of 7.4 $\pm$ 0.6 μM. Moreover, this compound showed selectivity for NTMT1 compared with PRMT1 and G9a. Even through the molecular structure of this compound remains to be validated, however it is the first non-substrate analogue inhibitor identified so far for NTMT1. It can be used as a lead for inhibitor design.

Lastly, our results suggest that NTMT2 is able to di- and tri-methylate those substrate peptides that have high binding affinity to NTMT1/2, such as GPKRRQ, GPRRRS, and PPKRIA (MePPKRIA). However, most peptides that can be di-methylated by NTMT1 were not being able to di-methylated by NTMT2. Our studies confirmed that NTMT2 can monomethylate all NTMT1 substrates. Furthermore, we discovered that it is capable of di-, tri-methylate some substrates with high binding affinity, which implies that NTMT2 is not a monomethylase.

4. Future direction

Even though the N-terminal methylation modification has been known for a long time, it began to stimulate increasing interest after the discovery of NTMT1 in 2010. Studies about this modification and its enzymes are still at very early stage. In our study, we have demonstrated that NTMT1 is adopting Bi-Bi mechanism while catalysis reaction and its substrates are methylated in a distributive fashion. Our structural basis study indicated a few important residues in substrate binding site and refined its substrate motif to X-P-K/R, where X can be any amino acid except D/E. However, there remains little explanation for the high binding affinity of peptides with X = Pro (PPKRIA), which is derived from mouse RCC1. Also the catalytic mechanism of NTMT1 still needs to be further elucidated. Structural characterization of our identified small molecule inhibitor remains to be clarified.

Compared to NTMT1, NTMT2 is a newly identified N-methyltransferse. It was reported as a monomethylase that plays a role to prime the substrates of NTMT1 to facilitate further methylation of substrates by NTMT1. However, our methylation progression assay indicates that NTMT2 is capable of di- and tri-methylation of some peptides that have high binding affinity to NTMT2. Future studies are needed to understand whether this is due to substrate specificity or the methylation rate difference between the two homologues.

Protein α-N-terminal methylation plays an important role in regulating protein-DNA interactions, mitotic division and DNA damage repair; however, little is known about the biological functions of this modification. Therefore, identification of proteins that recognize and bind to N-terminal methylated proteins would advance our understanding of its biological functions.

5. Experimental and methods

5.1 Materials and instruments

SAM, ThioGlo1, Tris, KCl, NaCl, TECP, $NH_4H_2PO_4$, TFA, dimethylformamide, α-Cyano-3-hydroxycinnamic acid, amino acids, and other chemicals and reagents were purchased from VWR, Fischer, Aldrich, EMD, Caliochem and Chemlmpex. NTMT1 (AD-003) clone was purchased from Addgene. SAHH clone was obtained from Dr. Raymond C. Trievel through a Materials Transfer Agreement.

Flexstation3 Multi-Mode Microplate Reader was utilized for NTMT1 kinetic characterization. And methylation progression assay was carried out via an Applied Biosystems Voyager matrix-assisted laser desorption/ionization time-of-flight mass spectrometer.

5.2 NTMT1 purification

NTMT1 preparation for kinetic studies

His-NTMT1 was expressed in *E. coli* BL21 (DE3) codon plus RIL cells in Terrific Broth medium in the presence of 50 g/mL kanamycin, using a pET28a-LIC expression vector that encodes a full-length NTMT1 (aminoacids1–222) with His_6 tag obtained from Addgene. Cells were grown at 37 °C to A_{600} of 1.5, induced by isopropyl

β-$_D$-1-thiogalactopyranoside (final concentration 1 mM), and incubated overnight at 15 °C. Cells were harvested by centrifugation at 5,000 rpm. Cell pellets were suspended in the lysis buffer containing 25 mM Tris-HCl buffer (pH 8.0) containing 300 mM NaCl and 10 mM imidazole, lysed by passing through a Microfluidizer (MicrofluidicsCorp.) at 20,000 p.s.i., and centrifuged at 15,000 rpm for 15 min at 4 °C. The supernatant was loaded on to a nickel-nitrilotriacetic acid agarose column. After removal of unbound protein by extensive washing with the lysis buffer, the protein was eluted with 25 mM Tris-HCl buffer (pH 8.0) containing 100 mM imidazole and 300 mM NaCl. Combined elution fractions were dialyzed in the dialysis buffer (25 mM Tris, pH7.5, 150 mM NaCl, 50 mM KCl) three times to provide His-NTMT1. The yield was 20 mg/L.

NTMT1 preparation for crystal structure and mutagenesis

The gene of human NTMT1 (2-223) was amplified and cloned into a modified pET28a-LIC vector to express NTMT1 with a 6× His-tag and a thrombin cleavage site at the N-terminus. The recombinant NTMT1 plasmid was transformed into *E. coli* BL21 (DE3) codon plus RIL strain for induced expression with 0.2 mM IPTG at 16 °C overnight. NTMT1 was purified by Ni^{2+}-affinity and anion-exchange chromatography, followed by further purification through Superdex™ 200 10/300 (GE Healthcare). The buffer for gel filtration contained 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM TECP. The peak fractions were collected and concentrated to 37 mg/mL for crystallization assay. The mutant proteins were purified using the same procedure as described above.

5.3 Peptide preparation

hRCC1-6 (SPKRIA, $[M+H]^+$ = 670.4359), hRCC1-9 (SPKRIAKRR, $[M+H]^+$ = 1110.7331), hRCC1-10 (SPKRIAKRRS, $[M+H]^+$ = 1197.7651), and hRCC1-12 (SPKRIAKRRSPP, $[M+H]^+$ = 1392.8546) were synthesized on Rink amide resin using standard Fmoc chemistry with a CEM Liberty microwave peptide synthesizer. Fmoc protection groups at the α-N-termini were removed by 20% (v/v) piperidine in N,N-dimethylformamide. MeRCC1-10 was synthesized based on literature.[103] The Me2-RCC1-10 peptide was synthesized as follows. To the deprotected RCC1-10 peptide on resin (0.1 mmol) in N,N-dimethylformamide (2 mL) was added formaldehyde (24 µL of 37% (w/v) solution, 0.24 mmol), HOAc (20 µL), and $NaBH_3CN$ (15 mg, 0.24 mmol). The mixture was placed on a shaker for 4 h. The resin was washed with N,N-dimethylformamide, and the reaction was repeated.

5.4 Fluorescence intensity *vs.* [GSH] correlation study

A three-fold serial dilution of GSH (0~30 µM) with 1x buffer solution (25 mM Tris-HCl, pH 7.5, 50 mM KCl) was conducted. The reaction was initiated by addition of 10 µL of GSH solution in different concentration to each well of 96-well microplate containing 10 µL of 10x reaction buffer (250 mM Tris-HCl, pH 7.5, 500 mM KCl), 1 µL ThioGlo1 (1.5 mM stock solution in DMSO) and 79 µL ddH_2O. Subsequently, 96-well microplate with testing sample was placed into FlexStation 3 microplate reader preset at 37 °C. The program was set as: kinetics, fluorescence bottom read, excitation of 370 nm, emission of 500 nm, duration 15 min with measurements at 15 seconds intervals.

After the fluorescence intensity was measured, the obtained raw data was processed to make a plot of Relative Fluorescence Units (RFU) as a function of GSH concentrations in Microsoft Office Excel scatter plot. The value for R^2 was calculated after linear regression.

5.5 Time-dependent studies

A reaction mixture of 0.2 μM NTMT1, 10 μM SAHH, 100 μM SAM, and 15 μM ThioGlo1 in buffer (25 mM Tris, 50 mM KCl, pH = 7.5) was premixed (90 μL in total) and transferred to a well of 96-well microplate. The reaction was initiated with 10 μL of 50 μM RCC1-10, and incubated in the FlexStation 3 at 37 °C for 5 min. The program of FlexStation 3 was set as: kinetics, fluorescence bottom read, excitation of 370 nm, emission of 500 nm, duration 15 min with measurements at 15 seconds intervals. And the reaction was monitored continuously for 15 min. The obtained RFU was converted to [SAH] using the standard curve obtained (RFU *vs.* [GSH]). Then, data was processed to make a plot of [SAH] as a function of time as a Microsoft Office Excel scatter plot. The value for R^2 was calculated after linear regression.

5.6 Concentration-dependent studies

A reaction mixture of 10 μM SAHH, 100 μM SAM, and 15 μM ThioGlo1 in 1x buffer solution (25 mM Tris-HCl, pH 7.5, 50 mM KCl) was premixed (89 μL in total) and transferred to each well of a 96-well microplate. NTMT1 was diluted with 1x buffer in

different concentrations: 40 µM, 30 µM, 20 µM, 10 µM, 5 µM. 1 µL of NTMT1 solution at varied concentration was added into the above reaction mixture. The reaction was next initiated with 10 µL of 50 µM of RCC1-10, incubated under 37 °C for 5 min, and monitored by the FlexStation3 microplate reader (kinetics, fluorescence bottom read, excitation of 370 nm, emission of 500 nm, duration 15 min with measurements at 15 seconds intervals). The obtained RFU was converted to [SAH] using the standard curve obtained (RFU *vs.* [GSH]). Then, data was processed to make a plot of [SAH] as a function of time in Microsoft Office Excel scatter plot. Initial rates were calculated and processed to make a plot of rate as a function of [NTMT1]. Then, data was processed to make a plot of [SAH] as a function of time as a Microsoft Office Excel scatter plot.

5.7 Steady state kinetic characterization of NTMT1 substrates

5.7.1 SAM

Serial dilution of SAM was carried out to a final concentration range (100 µM, 50 µM, 25 µM, 12.5 µM, 6.3 µM, 3.1 µM, 1.6 µM). A reaction mixture of 0.2 µM NTMT1, 10 µM SAHH, SAM, and 15 µM ThioGlo1 was premixed in the 1x reaction buffer (25 mM Tris, 50 mM KCl, pH 7.5) and 90 µL was transferred to each well of 96-well microplate and incubated at 37 °C for 5 min. The reaction was next initiated with 10 µL of 50 µM of RCC1-12, incubated under 37 °C for 5 min, and monitored by the FlexStation3 microplate reader (kinetics, fluorescence bottom read, excitation of 370 nm, emission of 500 nm, duration 15 min with measurements at 15 seconds intervals). The obtained data were processed as described before and subsequently fit into Michaelis-Menten

model using least square nonlinear regression regression with GraphPad Prism 5 (Version 5.04).

5.7.2 Peptide substrates

Each peptide was diluted to 10 mM using deionized water. Serial dilution was carried out to a final concentration range (64 µM, 32 µM, 16 µM, 8 µM, 4 µM, 2 µM, 1 µM, 0.5 µM, 0). A test run was carried out following the fluorescence assay as described previously to estimate the concentration range for kinetics studies. Based on the results of these test runs, specific serial dilutions tailored for each peptide were adjusted.

A reaction mixture of 0.2 µM NTMT1, 10 µM SAHH, 100 µM SAM, and 15 µM ThioGlo1 was premixed in the 1x reaction buffer (25 mM Tris, 50 mM KCl, pH 7.5) and 90 µL was transferred to each well of 96-well microplate and incubated at 37 °C for 5 min. The reaction was next initiated with addition of varied concentrations of RCC1-12 in different wells and monitored by the FlexStation3 microplate reader for 15 min. The result was processed by Microsoft Office Excel scatter plot. The obtained data were processed as described before and subsequently fit into Michaelis-Menten model using least square nonlinear regression regression with GraphPad Prism 5 (Version 5.04).

5.8 NTMT1 kinetic mechanism characterization

RCC1-10 was prepared in varied concentrations (20 µM, 10 µM, 5 µM, 2.5 µM, 1.25 µM, 0.625 µM) through serial dilution with deionized water. Four kinetic

experiments with varied SAM concentration (12.5 μM, 25 μM, 50 μM, 100 μM) were carried out in parallel. The detailed experimental procedures were as described above. Obtained initial velocities from each experiment were globally fitted into Lineweaver-Burk kinetic equation using least square nonlinear regression GraphPad Prism 5 software (Version 5.04).

For the SAM binding site, SAM was prepared in varied concentration (51.2 μM, 25.6 μM, 12.8 μM, 6.4 μM, 3.2 μM, 1.6 μM) through serial dilution with deionized water. Different RCC1-10 concentrations (1 μM, 2 μM, 4 μM, 8 μM) were used for each experiment. Similarly, the obtained initial velocity from each experiment was globally fit with the Lineweaver-Burk kinetic equation using least square nonlinear regression GraphPad Prism 5 software (Version 5.04).

5.9 MALDI-MS methylation progression study

NTMT1 methylation progression studies

The reaction buffer (20 mM Tris, 50 mM KCl, pH 7.5), RCC1-10 (10 μM), and NTMT1 (0.2 μM) were premixed, and incubated in a 30 °C water bath for 5 min. After incubation, the reaction was initiated by addition of 50 μM SAM. The reaction took 2 hours. Every 20 min, a 3 μL sample was withdrawn from the reaction mixture, and quenched in 1:1 ratio using quenching solution (20 mM $NH_4H_2PO_4$, 0.4% (v/v) TFA, in 1:1 ACN/water). 1 μL quenched samples were spotted onto MALDI plate along with α-Cyano-3-hydroxycinnamic acid, and examined by MALDI-MS.

Methylation progression studies of peptides of varied first residue with NTMT1 and NTMT2

A 36 µL reaction mixture was prepared, which is composed of: 2 µL NTMT1/2 (40 µM), 30 µL reaction buffer (20 µM Tris, 50 µM NaCl, pH = 7.5), and 4 µL peptide substrate (200 µM). This reaction mixture was incubated in 30 °C water bath for 5 minutes, and initiated with 4 µL SAM (400 µM). At each time point (0 min, 15 min, 30 min, 60 min, 120 min, overnight), 3 µL of reaction mixture was withdrawn and quenched in 1:1 ratio by quenching solution (20 mM $NH_4H_2PO_4$, 0.4% (v/v) TFA, in 1:1 ACN/water). Quenched samples were spotted onto MALDI plate along with α-cyano-3-hydroxycinnamic acid, and examined by MALDI-MS.

MALDI-MS based study of endogenous SAM

Buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.5) was mixed with 50 µM SAM and 1 mM peptide (SPKRIA, PPKRIA, RPKRIA, YPKRIA). Reaction mixtures were incubated at 30 °C for 2 h. After 2 h, the testing samples were withdrawn from each reaction mixtures were subsequently quenched in 1:1 ratio by the quenching solution (20 mM $NH_4H_2PO_4$, 0.4% (v/v) TFA, in 1:1 ACN/water). These quenched samples were spotted onto a MALDI plate along with α-Cyano-3-hydroxycinnamic acid, and examined by MALDI-MS.

5.10 Virtual screening and docking studies

By using the NTMT1 – SAH binary complex (PDB ID: 2EX4, 1.75Å) crystal structure model, a query was designed based on the predicted interactions between substrate and NTMT1. An aromatic region was defined complementary to Trp137. And a hydrophobic region was defined in the center of Val218, Leu211, and Ile37. The acid side chain of Asp181 was defined as an acceptor atom. Finally, a negative center was defined (surrounded by Asp181, Asp178, Asp168, and Ser163). Subsequently, a virtual screening was carried out by using designed query. Libraries used for virtual screening include NCI, Asinex, Chembridge, Maybridge, Otava, and SigmaAldric. In total, there were around 2,500,000 compounds were surveyed.

Hits obtained from virtual screening were docked into both binding sites of NTMT1. Results were shown as scores using HINT (**H**ydropathic **INT**eractions) score functions. Both SAM and peptide binding sites were scored. Docking scores of SAM binding site was used as a reference to compare with those of the peptide binding site. Molecules with high scores at the peptide binding site, and low scores at the SAM binding site were expected. This observation suggests that such hits have higher possibilities of targeting the peptide binding site.

5.11 Biochemical screening of small molecule compounds

5.11.1 Primary screening using SAHH-coupled fluorescence assay

The samples to be tested were diluted and prepared as 10 mM stock solutions in 1.5 mL amber vials. 1 µL (10 mM) of each sample was added into each well of microplate. Subsequently, 90 µL of reaction mixture was prepare for each well of a 96-well microplate. The reaction mixture is composed of 0.2 µM NTMT1, reaction buffer (20 mM Tris, 50 mM KCl, pH 7.5), 10 µM SAHH, 100 µM SAM (10 x K_m concentration) and 15 µM ThioGlo1. Reactions were initiated with 10 µL (50 µM) (1 x K_m concentration) of RCC1-10, and measured by the FlexStation3 microplate reader. The obtained raw data were processed by Microsoft Office Excel in scatter plots.

5.11.2 Secondary screening to remove false positive

Four parallel inhibition assays of each candidate were carried out, preparation procedure and methodology were as described previously. The first and second experiments have varied SAHH concentrations, 1 x SAHH and 2 x SAHH respectively. And the third and fourth experiments have 0.01% and 0.04% of triton. The obtained data were processed by Microsoft Office Excel in scatter plots.

5.11.3 IC_{50} studies

Three-fold serial dilution of each sample was carried out with DMSO, and 1 µL of each sample in different concentrations were added into a 96-well microplate. Reaction mixtures were prepared as described previously, and the reaction was initiated

with 10 uL (50 μM) RCC1-10, incubated under 37 °C, and read by FlexStation3 microplate reader. The obtained data was firstly processed by Microsoft Office Excel in scatter plots; and IC_{50} profiles were determined through SigmaPlot 5 using “response *vs*. Log conc.” of non-linear regression.

5.11.4 Selectivity studies

IC_{50} studies were conducted as described previously with NTMT1, PRMT5, and G9a, respectively. Obtained data were processed by Microsoft Office Excel in scatter plots, and the IC_{50} profiles were determined through SigmaPlot 5 “response *vs*. Log conc.” of non-linear regression.

References

1. Lothrop, A. P.; Torres, M. P.; Fuchs, S. M. Deciphering post-translational modification codes. *FEBS Lett.* **2013***, 587*, 1247-1257.

2. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2012***, 4*, 565-583.

3. Beltrao, P.; Bork, P.; Krogan, N. J.; van Noort, V. Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* **2013***, 9*, 714.

4. Strahl, B. D.; Allis, C. D. The language of covalent histone modifications. *Nature.* **2000***, 403*, 41-45.

5. Verdin, E.; Ott, M. 50 Years of Protein Acetylation: from Gene Regulation to Epigenetics, Metabolism and Beyond. *Nat. Rev. Mol. Cell Biol.* **2015***, 16*, 258-264.

6. Glozak, M. A.; Sengupta, N.; Zhang, X.; Seto, E. Acetylation and deacetylation of non-histone proteins. *Gene.* **2005***, 363*, 15-23.

7. PHILLIPS, D. M. The presence of acetyl groups of histones. *Biochem. J.* **1963***, 87*, 258-263.

8. ALLFREY, V. G.; FAULKNER, R.; MIRSKY, A. E. Acetylation and Methylation of Histones and their Possible Role in the Regulation of Rna Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1964***, 51*, 786-794.

9. Vidali, G.; Boffa, L. C.; Bradbury, E. M.; Allfrey, V. G. Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **1978***, 75*, 2239-2243.

10. Shi, Y.; Lee, J. S.; Galvin, K. M. Everything you have ever wanted to know about Yin Yang 1. *Biochim. Biophys. Acta.* **1997***, 1332*, F49-66.

11. Espinosa, J. M.; Emerson, B. M. Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed cofactor recruitment. *Mol. Cell.* **2001***, 8*, 57-69.

12. Bustin, M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell. Biol.* **1999***, 19*, 5237-5246.

13. Bustin, M.; Reeves, R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* **1996***, 54*, 35-100.

14. Jeong, J. W.; Bae, M. K.; Ahn, M. Y.; Kim, S. H.; Sohn, T. K.; Bae, M. H.; Yoo, M. A.; Song, E. J.; Lee, K. J.; Kim, K. W. Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation. *Cell.* **2002***, 111*, 709-720.

15. Kawai, H.; Li, H.; Avraham, S.; Jiang, S.; Avraham, H. K. Overexpression of histone deacetylase HDAC1 modulates breast cancer progression by negative regulation of estrogen receptor alpha. *Int. J. Cancer.* **2003***, 107*, 353-358.

16. Turner, B. M.; Birley, A. J.; Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell.* **1992***, 69*, 375-384.

17. Turner, B. M.; Fellows, G. Specific antibodies reveal ordered and cell-cycle-related use of histone-H4 acetylation sites in mammalian cells. *Eur. J. Biochem.* **1989***, 179*, 131-139.

18. Davie, J. R. Inhibition of histone deacetylase activity by butyrate. *J. Nutr.* **2003***, 133*, 2485S-2493S.

19. Johnson, L. N. The regulation of protein phosphorylation. *Biochem. Soc. Trans.* **2009***, 37*, 627-641.

20. Welburn, J. P.; Tucker, J. A.; Johnson, T.; Lindert, L.; Morgan, M.; Willis, A.; Noble, M. E.; Endicott, J. A. How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. *J. Biol. Chem.* **2007***, 282*, 3173-3181.

21. FISCHER, E. H.; KREBS, E. G. Conversion of phosphorylase b to phosphorylase a in muscle extracts. *J. Biol. Chem.* **1955***, 216*, 121-132.

22. Walsh, D. A.; Perkins, J. P.; Krebs, E. G. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *J. Biol. Chem.* **1968***, 243*, 3763-3765.

23. Johnson, L. N.; Barford, D. The effects of phosphorylation on the structure and function of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1993***, 22*, 199-232.

24. Johnson, L. N.; Lewis, R. J. Structural basis for control by phosphorylation. *Chem. Rev.* **2001***, 101*, 2209-2242.

25. Mandell, D. J.; Chorny, I.; Groban, E. S.; Wong, S. E.; Levine, E.; Rapp, C. S.; Jacobson, M. P. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *J. Am. Chem. Soc.* **2007***, 129*, 820-827.

26. Ferrarese, A.; Marin, O.; Bustos, V. H.; Venerando, A.; Antonelli, M.; Allende, J. E.; Pinna, L. A. Chemical dissection of the APC Repeat 3 multistep phosphorylation by the concerted action of protein kinases CK1 and GSK3. *Biochemistry.* **2007***, 46*, 11902-11910.

27. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science.* **2002***, 298*, 1912-1934.

28. Roskoski, R.,Jr Src protein-tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* **2004***, 324*, 1155-1164.

29. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell.* **2006***, 127*, 635-648.

30. Hurley, J. H.; Dean, A. M.; Thorsness, P. E.; Koshland, D. E.,Jr; Stroud, R. M. Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme. *J. Biol. Chem.* **1990***, 265*, 3599-3602.

31. Russo, A. A.; Jeffrey, P. D.; Pavletich, N. P. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol.* **1996***, 3*, 696-700.

32. Paik, W. K.; Paik, D. C.; Kim, S. Historical review: the field of protein methylation. *Trends Biochem. Sci.* **2007***, 32*, 146-152.

33. Schubert, H. L.; Blumenthal, R. M.; Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **2003***, 28*, 329-335.

34. Cantoni, G. L. Biological methylation: selected aspects. *Annu. Rev. Biochem.* **1975***, 44*, 435-451.

35. Paik, W. K.; Kim, S. Protein methylase I. Purification and properties of the enzyme. *J. Biol. Chem.* **1968***, 243*, 2108-2114.

36. Nomoto, M.; Kyogoku, Y.; Iwai, K. N-Trimethylalanine, a novel blocked N-terminal residue of Tetrahymena histone H2B. *J. Biochem.* **1982***, 92*, 1675-1678.

37. Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem.* **2012***, 13*, 2642-2655.

38. Ohsawa, N.; Tsujita, M.; Morikawa, S.; Itoh, N. Purification and characterization of a monohalomethane-producing enzyme S-adenosyl-L-methionine: halide ion methyltransferase from a marine microalga, Pavlova pinguis. *Biosci. Biotechnol. Biochem.* **2001***, 65*, 2397-2404.

39. Attieh, J. M.; Hanson, A. D.; Saini, H. S. Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in Brassica oleracea. *J. Biol. Chem.* **1995***, 270*, 9250-9257.

40. Patnaik, D.; Chin, H. G.; Esteve, P. O.; Benner, J.; Jacobsen, S. E.; Pradhan, S. Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. *J. Biol. Chem.* **2004***, 279*, 53248-53258.

41. Richardson, S. L.; Mao, Y.; Zhang, G.; Hanjra, P.; Peterson, D. L.; Huang, R. Kinetic mechanism of protein N-terminal methyltransferase 1. *J. Biol. Chem.* **2015***, 290*, 11601-11610.

42. Obianyo, O.; Thompson, P. R. Kinetic mechanism of protein arginine methyltransferase 6 (PRMT6). *J. Biol. Chem.* **2012***, 287*, 6062-6071.

43. Obianyo, O.; Osborne, T. C.; Thompson, P. R. Kinetic mechanism of Protein Arginine Methyltransferase 1. *Biochemistry.* **2008***, 47*, 10420-10427.

44. Bedford, M. T.; Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell.* **2009***, 33*, 1-13.

45. Zurita-Lopez, C. I.; Sandberg, T.; Kelly, R.; Clarke, S. G. Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming omega-NG-monomethylated arginine residues. *J. Biol. Chem.* **2012***, 287*, 7859-7870.

46. Di Lorenzo, A.; Bedford, M. T. Histone arginine methylation. *FEBS Lett.* **2011***, 585*, 2024-2031.

47. Katz, J. E.; Dlakic, M.; Clarke, S. Automated identification of putative methyltransferases from genomic open reading frames. *Mol. Cell. Proteomics.* **2003***, 2*, 525-540.

48. Kim, J. H.; Yoo, B. C.; Yang, W. S.; Kim, E.; Hong, S.; Cho, J. Y. The Role of Protein Arginine Methyltransferases in Inflammatory Responses. *Mediators Inflamm.* **2016***, 2016*, 4028353.

49. Lee, J. H.; Cook, J. R.; Yang, Z. H.; Mirochnitchenko, O.; Gunderson, S. I.; Felix, A. M.; Herth, N.; Hoffmann, R.; Pestka, S. PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *J. Biol. Chem.* **2005***, 280*, 3656-3664.

50. Migliori, V.; Muller, J.; Phalke, S.; Low, D.; Bezzi, M.; Mok, W. C.; Sahu, S. K.; Gunaratne, J.; Capasso, P.; Bassi, C.; Cecatiello, V.; De Marco, A.; Blackstock, W.; Kuznetsov, V.; Amati, B.; Mapelli, M.; Guccione, E. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **2012***, 19*, 136-144.

51. Stopa, N.; Krebs, J. E.; Shechter, D. The PRMT5 arginine methyltransferase: many roles in development, cancer and beyond. *Cell Mol. Life Sci.* **2015***, 72*, 2041-2059.

52. Wang, M.; Xu, R. M.; Thompson, P. R. Substrate specificity, processivity, and kinetic mechanism of protein arginine methyltransferase 5. *Biochemistry.* **2013***, 52*, 5430-5440.

53. Qian, C.; Zhou, M. M. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol. Life Sci.* **2006***, 63*, 2755-2763.

54. Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **2005***, 6*, 227.

55. Jenuwein, T.; Laible, G.; Dorn, R.; Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol. Life Sci.* **1998***, 54*, 80-93.

56. Tschiersch, B.; Hofmann, A.; Krauss, V.; Dorn, R.; Korge, G.; Reuter, G. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **1994***, 13*, 3822-3831.

57. Sims, R. J.,3rd; Nishioka, K.; Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **2003***, 19*, 629-639.

58. Feng, Q.; Wang, H.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **2002***, 12*, 1052-1058.

59. Wilson, J. R.; Jing, C.; Walker, P. A.; Martin, S. R.; Howell, S. A.; Blackburn, G. M.; Gamblin, S. J.; Xiao, B. Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell.* **2002***, 111*, 105-115.

60. Adachi, Y.; Pavlakis, G. N.; Copeland, T. D. Identification and characterization of SET, a nuclear phosphoprotein encoded by the translocation break point in acute undifferentiated leukemia. *J. Biol. Chem.* **1994***, 269*, 2258-2262.

61. Adachi, Y.; Pavlakis, G. N.; Copeland, T. D. Identification of in vivo phosphorylation sites of SET, a nuclear phosphoprotein encoded by the translocation breakpoint in acute undifferentiated leukemia. *FEBS Lett.* **1994***, 340*, 231-235.

62. AMBLER, R. P.; REES, M. W. Epsilon-N-Methyl-lysine in bacterial flagellar protein. *Nature.* **1959***, 184*, 56-57.

63. Martin, C.; Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **2005***, 6*, 838-849.

64. Ng, H. H.; Feng, Q.; Wang, H.; Erdjument-Bromage, H.; Tempst, P.; Zhang, Y.; Struhl, K. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* **2002***, 16*, 1518-1527.

65. Trievel, R. C. Structure and function of histone methyltransferases. *Crit. Rev. Eukaryot. Gene Expr.* **2004***, 14*, 147-169.

66. Wittmann-Liebold, B.; Pannenbecker, R. Primary structure of protein L33 from the large subunit of the Escherichia coli ribosome. *FEBS Lett.* **1976***, 68*, 115-118.

67. Alix, J. H.; Hayes, D.; Lontie, J. F.; Colson, C.; Glatigny, A.; Lederer, F. Methylated amino acids in ribosomal proteins from Escherichia coli treated with ethionine and from a mutant lacking methylation of protein L11. *Biochimie.* **1979***, 61*, 671-679.

68. Brosius, J.; Chen, R. The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett.* **1976***, 68*, 105-109.

69. Chang, C. N.; Schwartz, M.; Chang, F. N. Identification and characterization of a new methylated amino acid in ribosomal protein L33 of Escherichia coli. *Biochem. Biophys. Res. Commun.* **1976***, 73*, 233-239.

70. Stock, A.; Clarke, S.; Clarke, C.; Stock, J. N-terminal methylation of proteins: structure, function and specificity. *FEBS Lett.* **1987***, 220*, 8-14.

71. Stock, A.; Schaeffer, E.; Koshland, D. E.,Jr; Stock, J. A second type of protein methylation reaction in bacterial chemotaxis. *J. Biol. Chem.* **1987***, 262*, 8011-8014.

72. Henry, G. D.; Dalgarno, D. C.; Marcus, G.; Scott, M.; Levine, B. A.; Trayer, I. P. The occurrence of alpha-N-trimethylalanine as the N-terminal amino acid of some myosin light chains. *FEBS Lett.* **1982***, 144*, 11-15.

73. Webb, K. J.; Lipson, R. S.; Al-Hadid, Q.; Whitelegge, J. P.; Clarke, S. G. Identification of protein N-terminal methyltransferases in yeast and humans. *Biochemistry.* **2010***, 49*, 5225-5235.

74. Alamgir, M.; Eroukova, V.; Jessulat, M.; Xu, J.; Golshani, A. Chemical-genetic profile analysis in yeast suggests that a previously uncharacterized open reading frame, YBR261C, affects protein synthesis. *BMC Genomics.* **2008***, 9*, 583-2164-9-583.

75. Tooley, C. E.; Petkowski, J. J.; Muratore-Schroeder, T. L.; Balsbaugh, J. L.; Shabanowitz, J.; Sabat, M.; Minor, W.; Hunt, D. F.; Macara, I. G. NRMT is an alpha-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature.* **2010***, 466*, 1125-1128.

76. Dannenberg, J. H.; te Riele, H. P. The retinoblastoma gene family in cell cycle regulation and suppression of tumorigenesis. *Results Probl. Cell Differ.* **2006***, 42*, 183-225.

77. Giacinti, C.; Giordano, A. RB and cell cycle progression. *Oncogene.* **2006***, 25*, 5220-5227.

78. Petkowski, J. J.; Bonsignore, L. A.; Tooley, J. G.; Wilkey, D. W.; Merchant, M. L.; Macara, I. G.; Schaner Tooley, C. E. NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J.* **2013***, 456*, 453-462.

79. Petkowski, J. J.; Schaner Tooley, C. E.; Anderson, L. C.; Shumilin, I. A.; Balsbaugh, J. L.; Shabanowitz, J.; Hunt, D. F.; Minor, W.; Macara, I. G. Substrate specificity of mammalian N-terminal alpha-amino methyltransferase NRMT. *Biochemistry.* **2012***, 51*, 5942-5950.

80. Dai, X.; Otake, K.; You, C.; Cai, Q.; Wang, Z.; Masumoto, H.; Wang, Y. Identification of novel N-methylation modification of CENP-B that regulates its binding to the centromeric DNA. *J. Proteome Res.* **2013***, 12*, 4167-4175.

81. Tanaka, Y.; Nureki, O.; Kurumizaka, H.; Fukai, S.; Kawaguchi, S.; Ikuta, M.; Iwahara, J.; Okazaki, T.; Yokoyama, S. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J.* **2001***, 20*, 6612-6618.

82. Howman, E. V.; Fowler, K. J.; Newson, A. J.; Redward, S.; MacDonald, A. C.; Kalitsis, P.; Choo, K. H. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl. Acad. Sci. U. S. A.* **2000***, 97*, 1148-1153.

83. Sullivan, K. F. A solid foundation: functional specialization of centromeric chromatin. *Curr. Opin. Genet. Dev.* **2001***, 11*, 182-188.

84. Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal alpha-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **2007***, 9*, 596-603.

85. Moore, W.; Zhang, C.; Clarke, P. R. Targeting of RCC1 to chromosomes is required for proper mitotic spindle assembly in human cells. *Curr. Biol.* **2002***, 12*, 1442-1447.

86. Li, H. Y.; Wirtz, D.; Zheng, Y. A mechanism of coupling RCC1 mobility to RanGTP production on the chromatin in vivo. *J. Cell Biol.* **2003***, 160*, 635-644.

87. Nemergut, M. E.; Mizzen, C. A.; Stukenberg, T.; Allis, C. D.; Macara, I. G. Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. *Science.* **2001***, 292*, 1540-1543.

88. Bischoff, F. R.; Ponstingl, H. Catalysis of guanine nucleotide exchange of Ran by RCC1 and stimulation of hydrolysis of Ran-bound GTP by Ran-GAP1. *Methods Enzymol.* **1995***, 257*, 135-144.

89. Bischoff, F. R.; Ponstingl, H. Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. *Nature.* **1991***, 354*, 80-82.

90. Bilbao-Cortes, D.; Hetzer, M.; Langst, G.; Becker, P. B.; Mattaj, I. W. Ran binds to chromatin by two distinct mechanisms. *Curr. Biol.* **2002***, 12*, 1151-1156.

91. Goshima, G.; Kiyomitsu, T.; Yoda, K.; Yanagida, M. Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. *J. Cell Biol.* **2003***, 160*, 25-39.

92. Okada, T.; Ohzeki, J.; Nakano, M.; Yoda, K.; Brinkley, W. R.; Larionov, V.; Masumoto, H. CENP-B controls centromere formation depending on the chromatin context. *Cell.* **2007***, 131*, 1287-1300.

93. Nakagawa, H.; Lee, J. K.; Hurwitz, J.; Allshire, R. C.; Nakayama, J.; Grewal, S. I.; Tanaka, K.; Murakami, Y. Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. *Genes Dev.* **2002***, 16*, 1766-1778.

94. Sullivan, K. F.; Glass, C. A. CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. *Chromosoma.* **1991***, 100*, 360-370.

95. Tachiwana, H.; Kagawa, W.; Kurumizaka, H. Comparison between the CENP-A and histone H3 structures in nucleosomes. *Nucleus.* **2012***, 3*, 6-11.

96. Yeh, J. I.; Levine, A. S.; Du, S.; Chinte, U.; Ghodke, H.; Wang, H.; Shi, H.; Hsieh, C. L.; Conway, J. F.; Van Houten, B.; Rapic-Otrin, V. Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* **2012***, 109*, E2737-46.

97. Keeney, S.; Chang, G. J.; Linn, S. Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *J. Biol. Chem.* **1993***, 268*, 21293-21300.

98. Cai, Q.; Fu, L.; Wang, Z.; Gan, N.; Dai, X.; Wang, Y. alpha-N-methylation of damaged DNA-binding protein 2 (DDB2) and its function in nucleotide excision repair. *J. Biol. Chem.* **2014***, 289*, 16046-16056.

99. Moser, J.; Volker, M.; Kool, H.; Alekseev, S.; Vrieling, H.; Yasui, A.; van Zeeland, A. A.; Mullenders, L. H. The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair (Amst).* **2005***, 4*, 571-582.

100. Dowden, J.; Pike, R. A.; Parry, R. V.; Hong, W.; Muhsen, U. A.; Ward, S. G. Small molecule inhibitors that discriminate between protein arginine N-methyltransferases PRMT1 and CARM1. *Org. Biomol. Chem.* **2011***, 9*, 7814-7821.

101. Zhang, G.; Richardson, S. L.; Mao, Y.; Huang, R. Design, synthesis, and kinetic analysis of potent protein N-terminal methyltransferase 1 inhibitors. *Org. Biomol. Chem.* **2015***, 13*, 4149-4154.

102. Dong, C.; Mao, Y.; Tempel, W.; Qin, S.; Li, L.; Loppnau, P.; Huang, R.; Min, J. Structural basis for substrate recognition by the human N-terminal methyltransferase 1. *Genes Dev.* **2015***, 29*, 2343-2348.

103. Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **2003**, *112*, 711-723.

104. Borch, F.; Bernstein, D.; Durst, D. Cyanohydridoborate anion as a selective reducing agent. *J. Am. Chem. Soc.* **1971***, 93*, 2897-2904.

Appendix

Table 1. Molecular mass of peptides of varied methylation states

Mass	SPKRIA	PPKRIA	YPKRIA	RPKRIA	DPKRIA
M+H	671	681	747	740	699
M+Na	693	703	769	762	721
M+K	709	719	785	778	737
M+Me+H	685	695	761	754	713
M+Me+Na	707	717	783	776	735
M+Me+K	723	733	799	792	751
M+2Me+H	699	709	775	768	727
M+2Me+Na	721		797	790	749
M+2Me+K	737		813	806	765
M+3Me	683		789	782	741

Mass	WPKRIA	QPKRIA	GPKRRQ	GPRRRS	LPKRIA	NPKRIA
M+H	770	712	740	727	697	698
M+Na	792	734	762	749	719	720
M+K	808	750	778	765	735	736
M+Me+H	784	726	754	741	711	712
M+Me+Na	806	748	776	763	733	734
M+Me+K	822	764	792	779	749	750
M+2Me+H	798	740	768	755	725	726
M+2Me+Na	820	762	790	777	747	748
M+2Me+K	836	778	806	793	763	764
M+3Me	812	754	782	769	739	740

PPKRIA – NTMT1

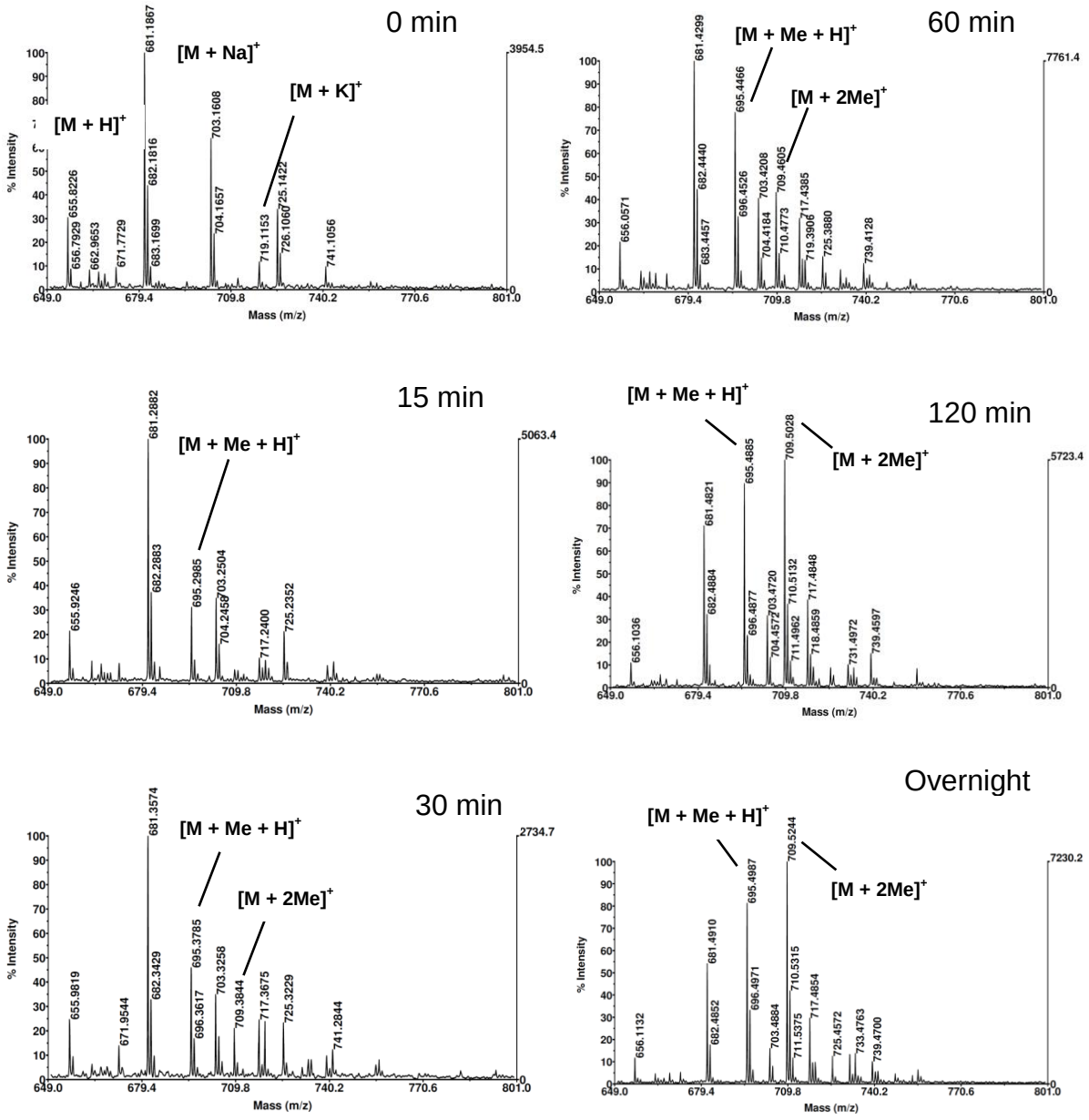


Figure 1. Methylation progression of PPKRIA with NTMT1.

PPKRIA – NTMT2

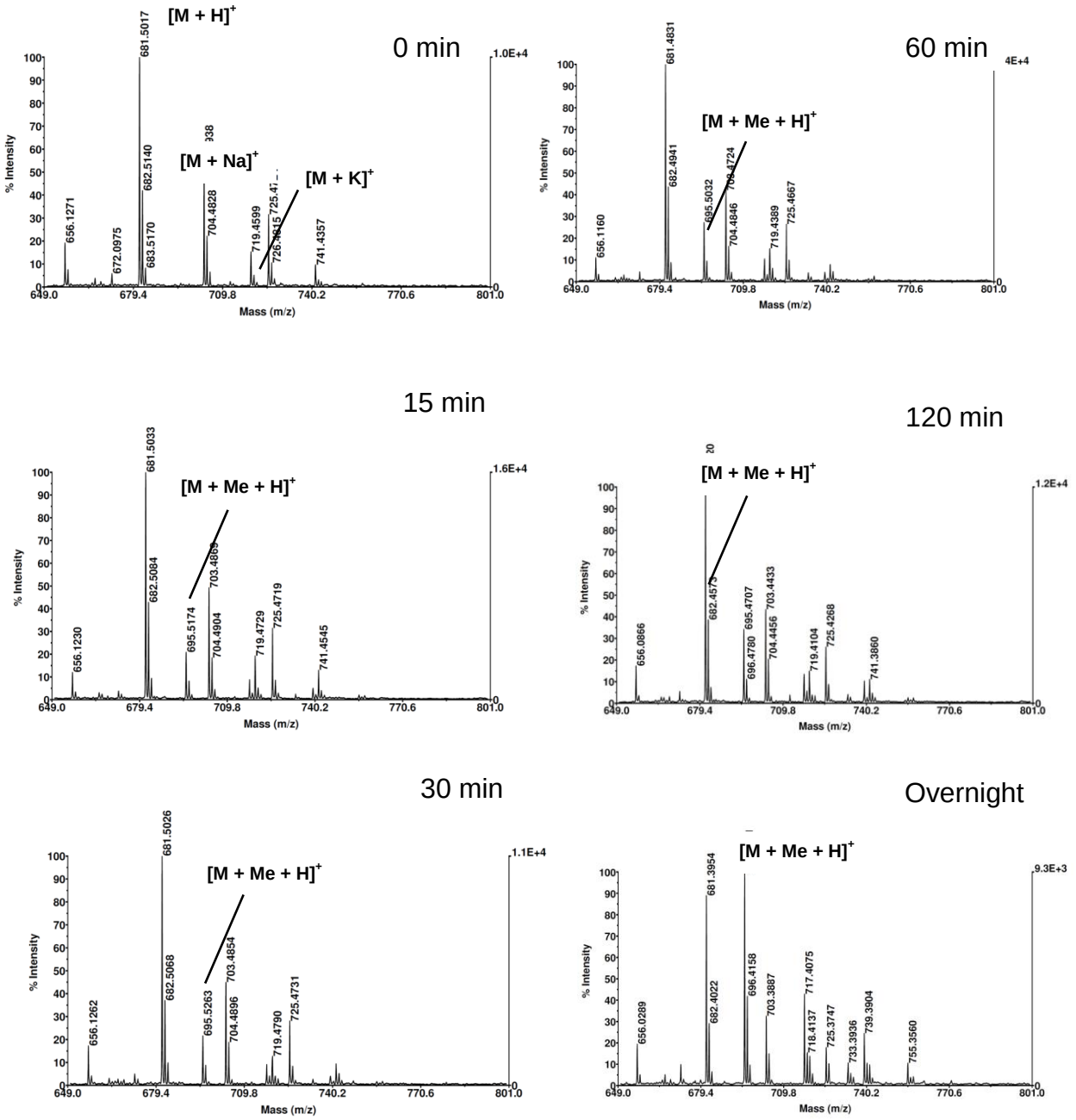


Figure 2. Methylation progression of PPKRIA with NTMT2.

GPKRRQ – NTMT1

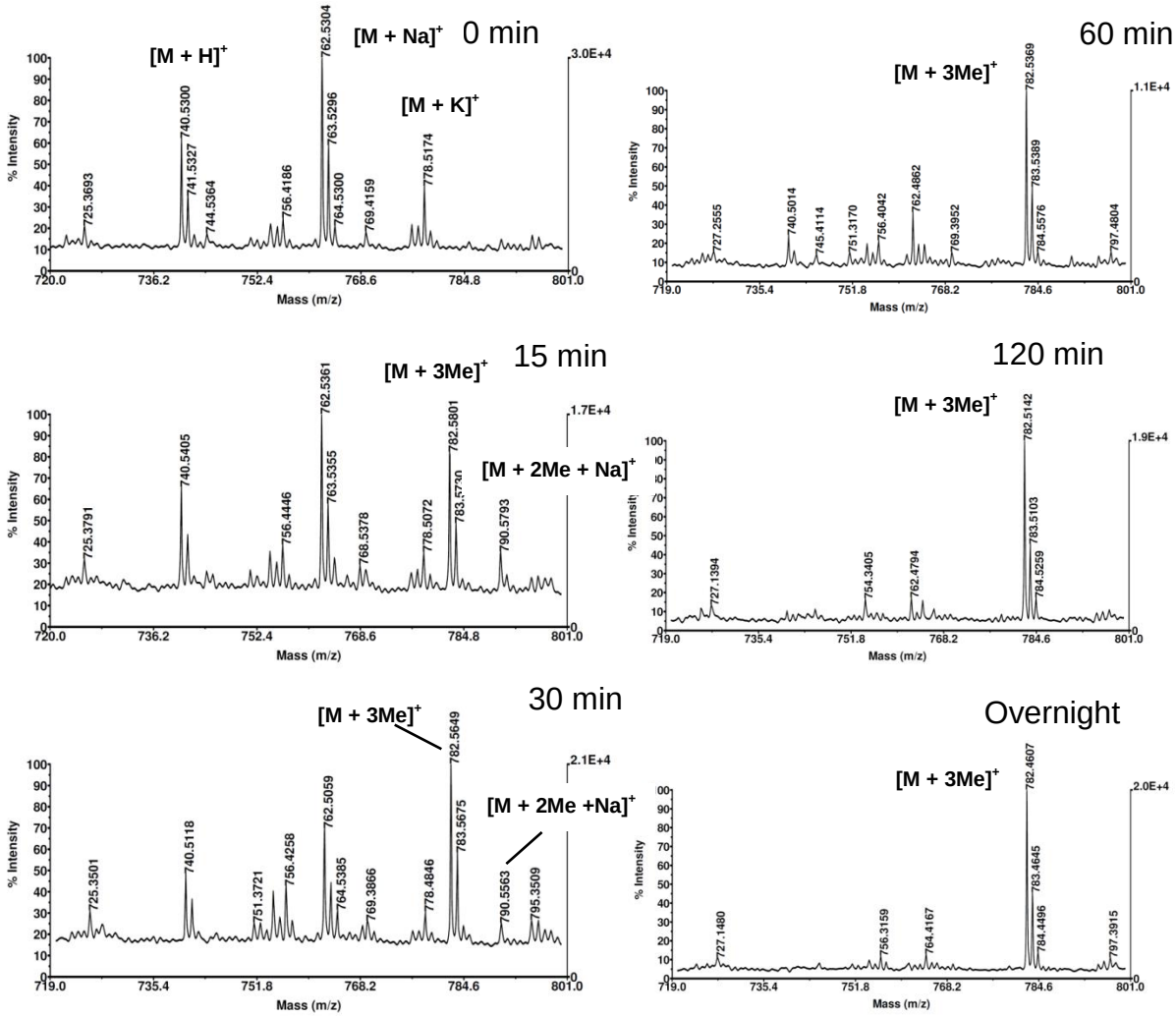


Figure 3. Methylation progression of GPKRRQ with NTMT1.

GPKRRQ – NTMT2

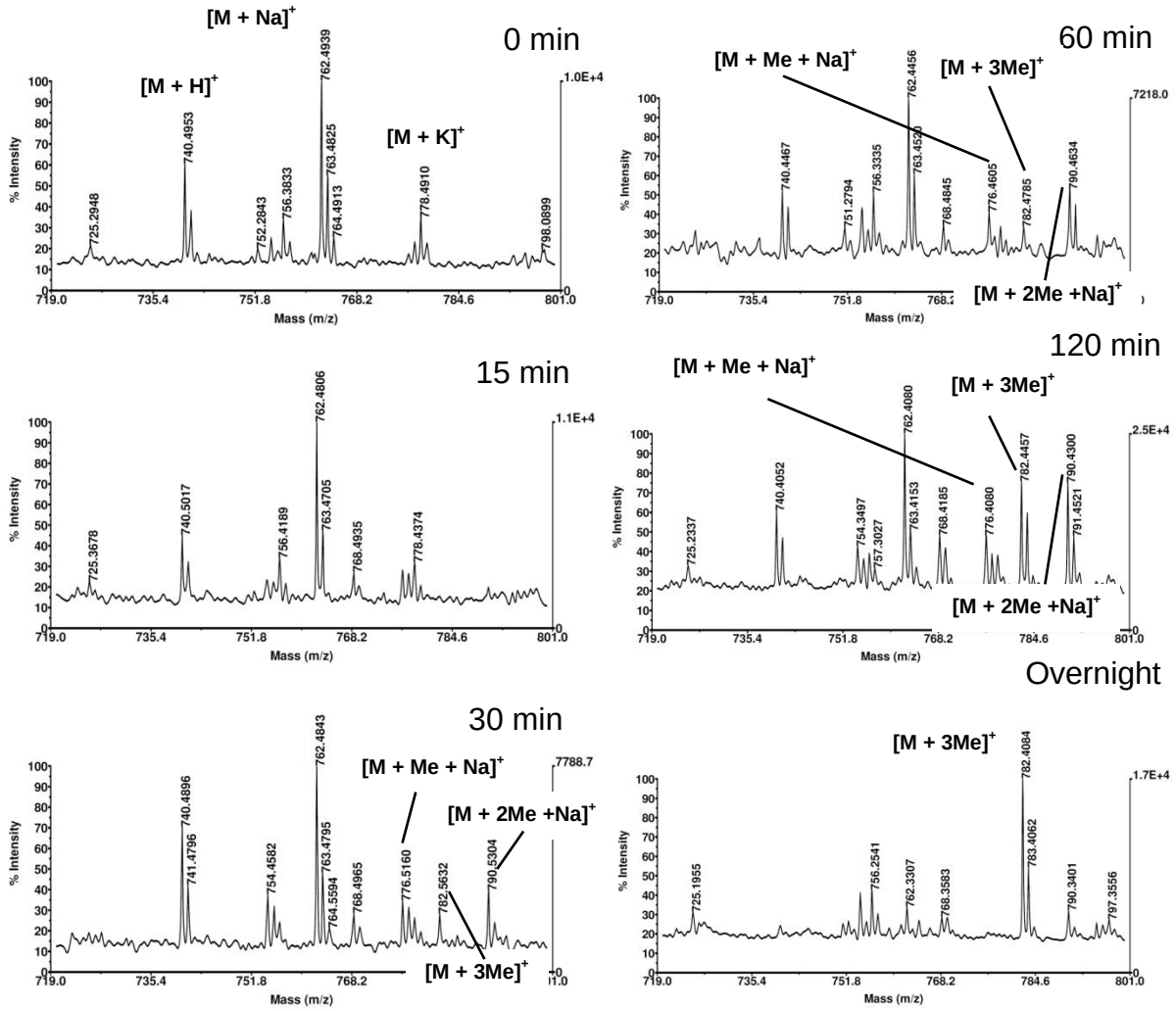


Figure 4. Methylation progression of GPKRRQ with NTMT2.

GPRRRS – NTMT1

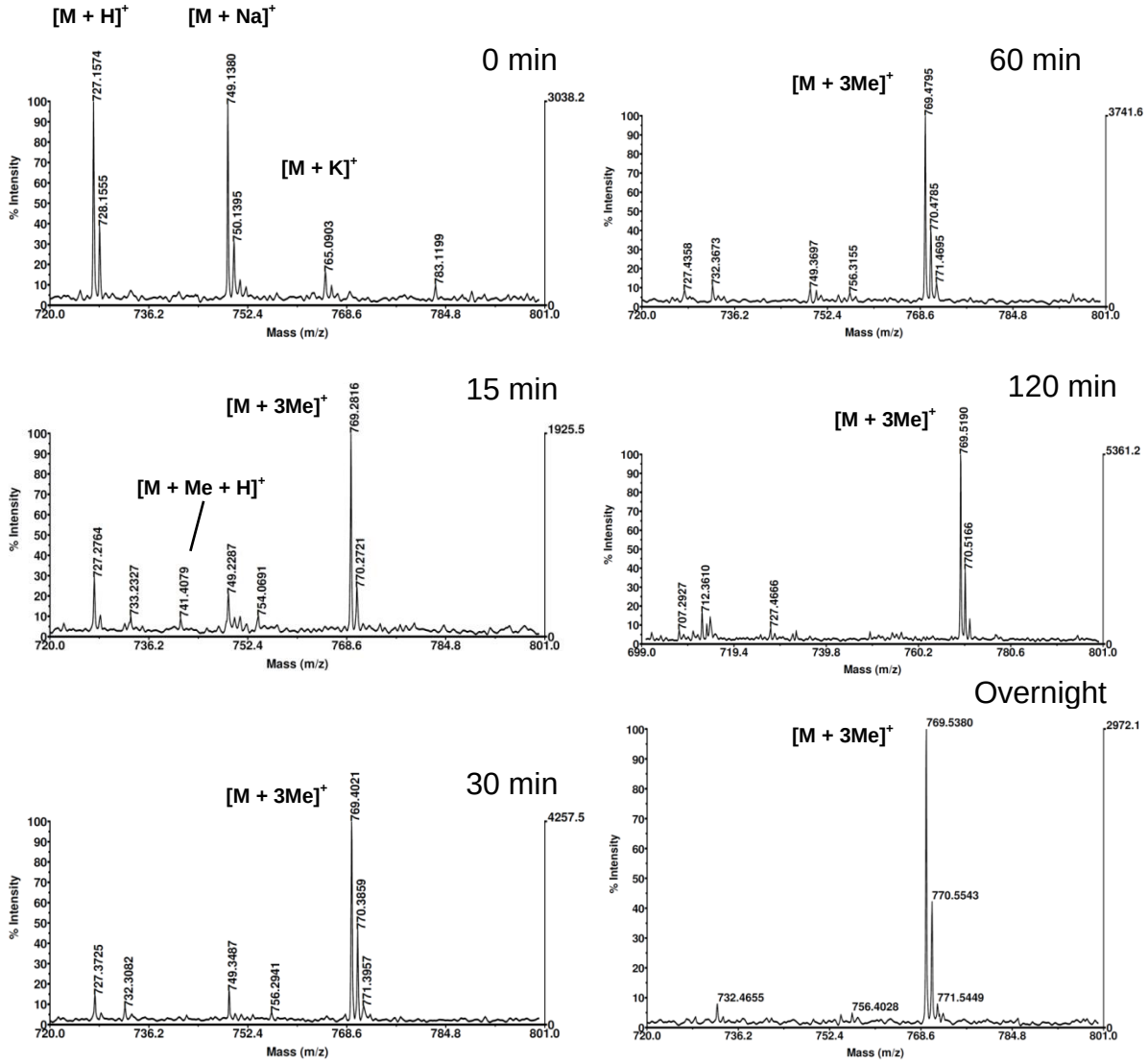


Figure 5. Methylation progression of GPRRRS with NTMT1.

GPRRRS – NTMT2

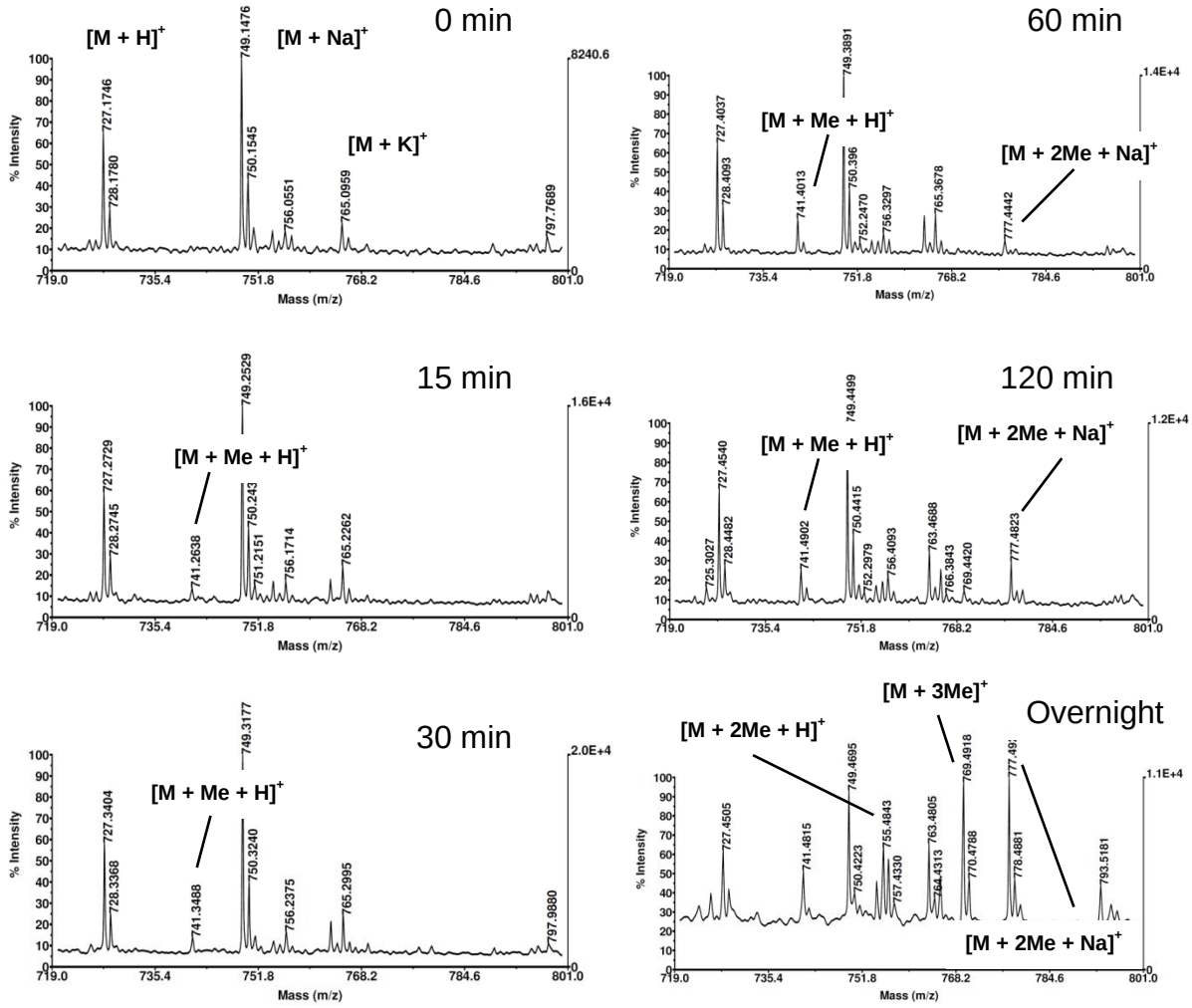


Figure 6. Methylation progression of GPRRRS with NTMT2.

SPKRIA – NTMT1

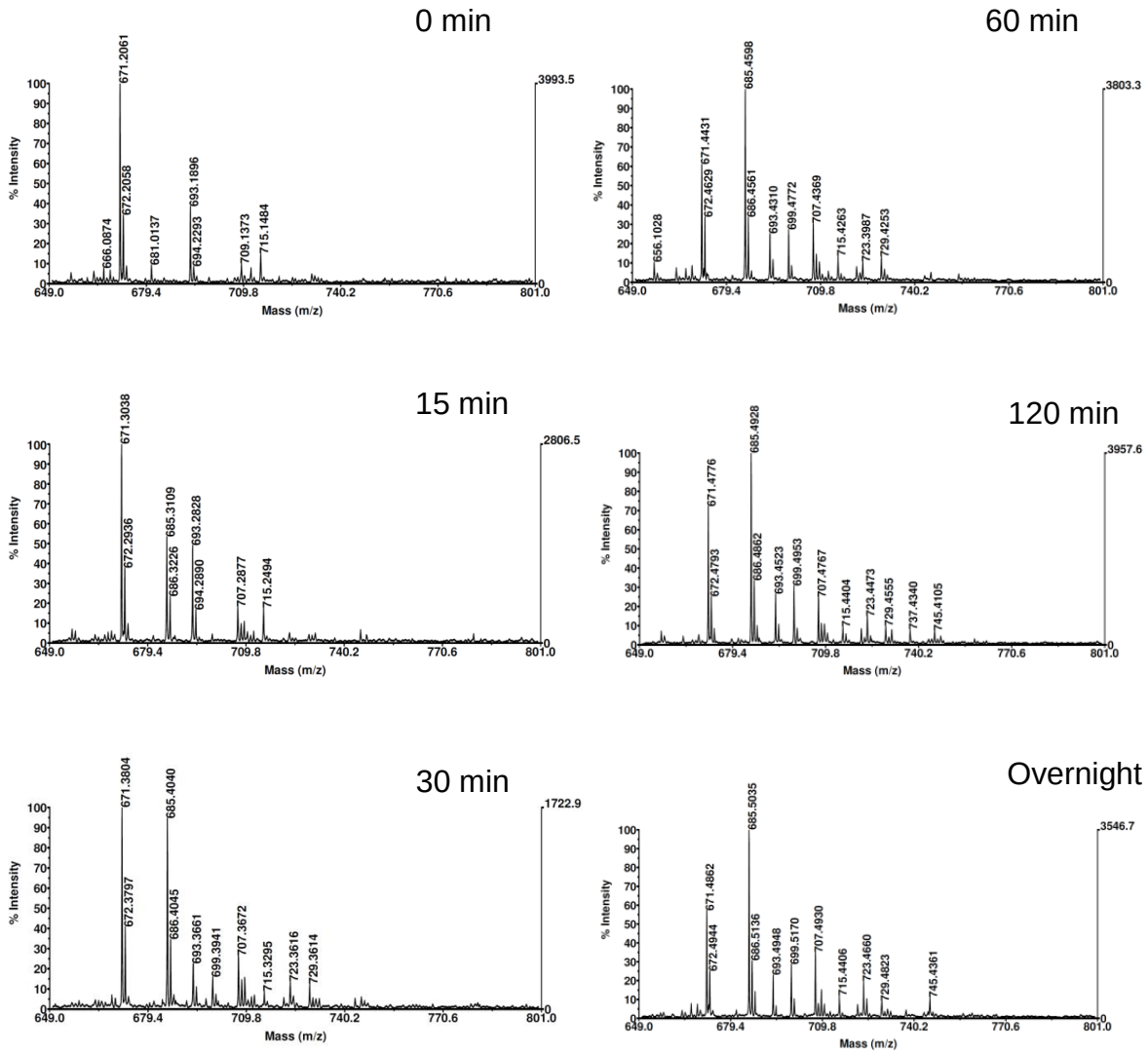


Figure 7. Methylation progression of SPKRIA with NTMT1.

SPKRIA – NTMT2

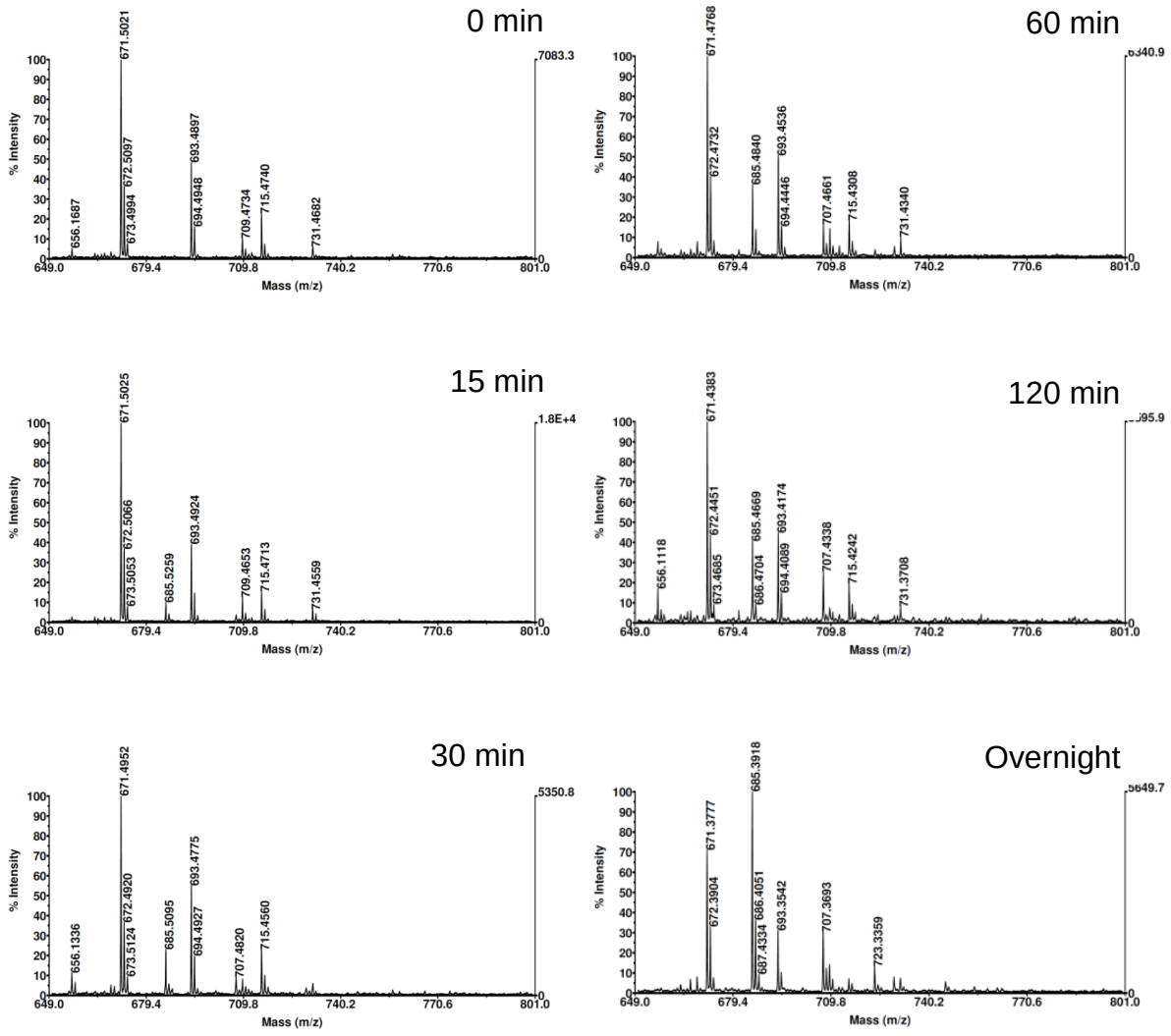


Figure 8. Methylation progression of SPKRIA with NTMT2.

MeSPKRIA – NTMT1

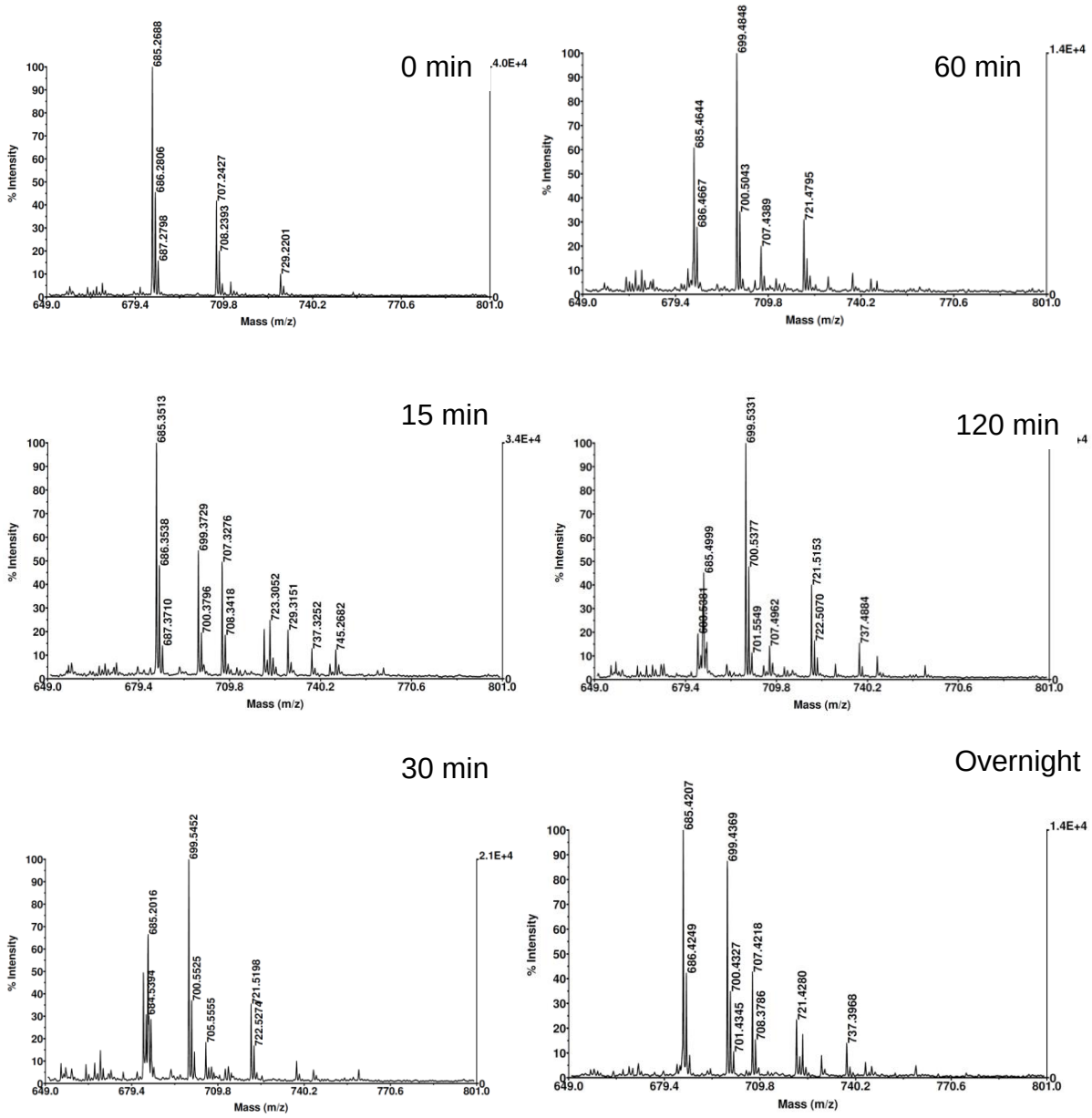


Figure 9. Methylation progression of MeSPKRIA with NTMT1.

MeSPKRIA – NTMT2

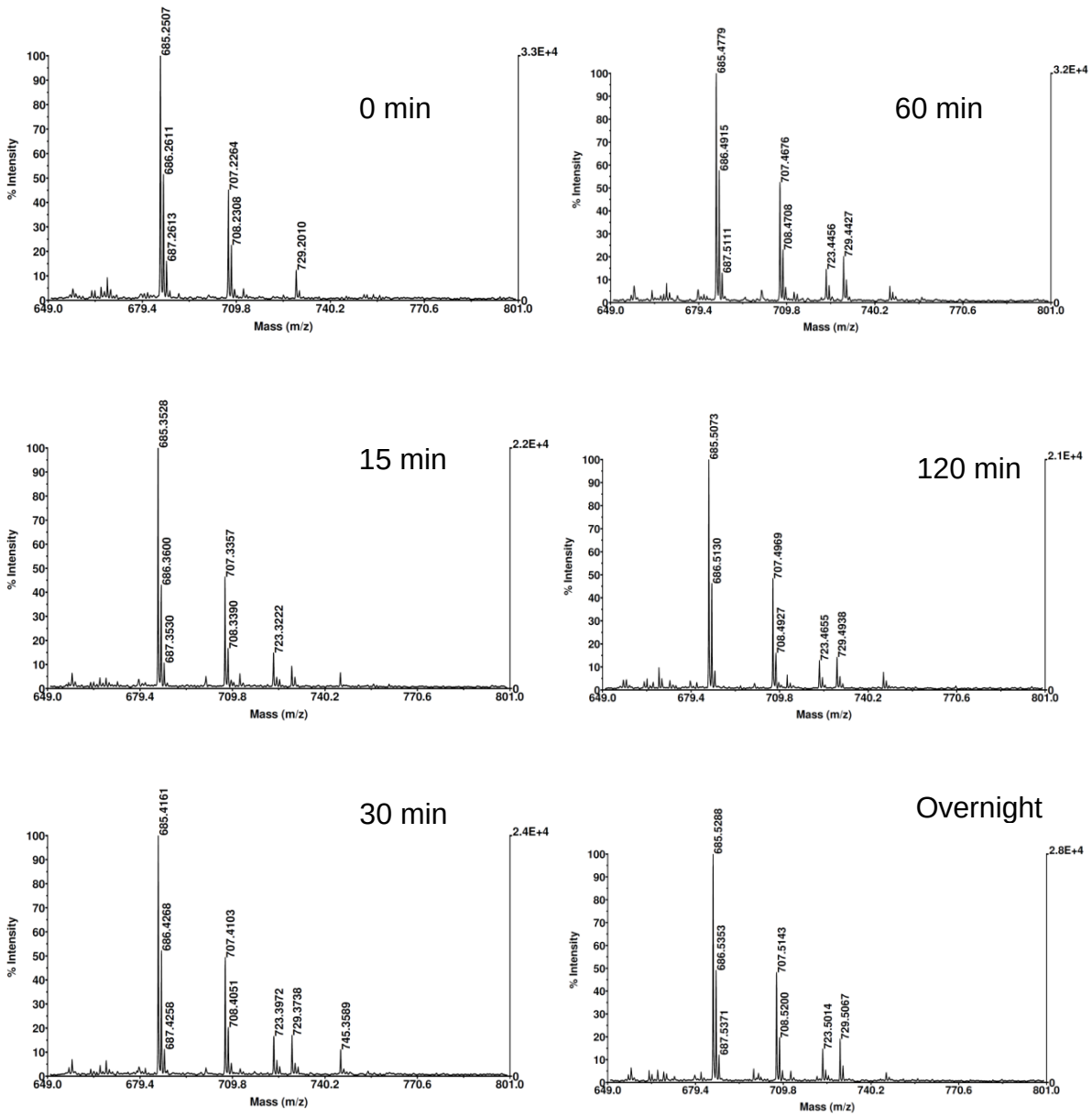


Figure 10. Methylation progression of MeSPKRIA with NTMT2.

RPKRIA – NTMT1

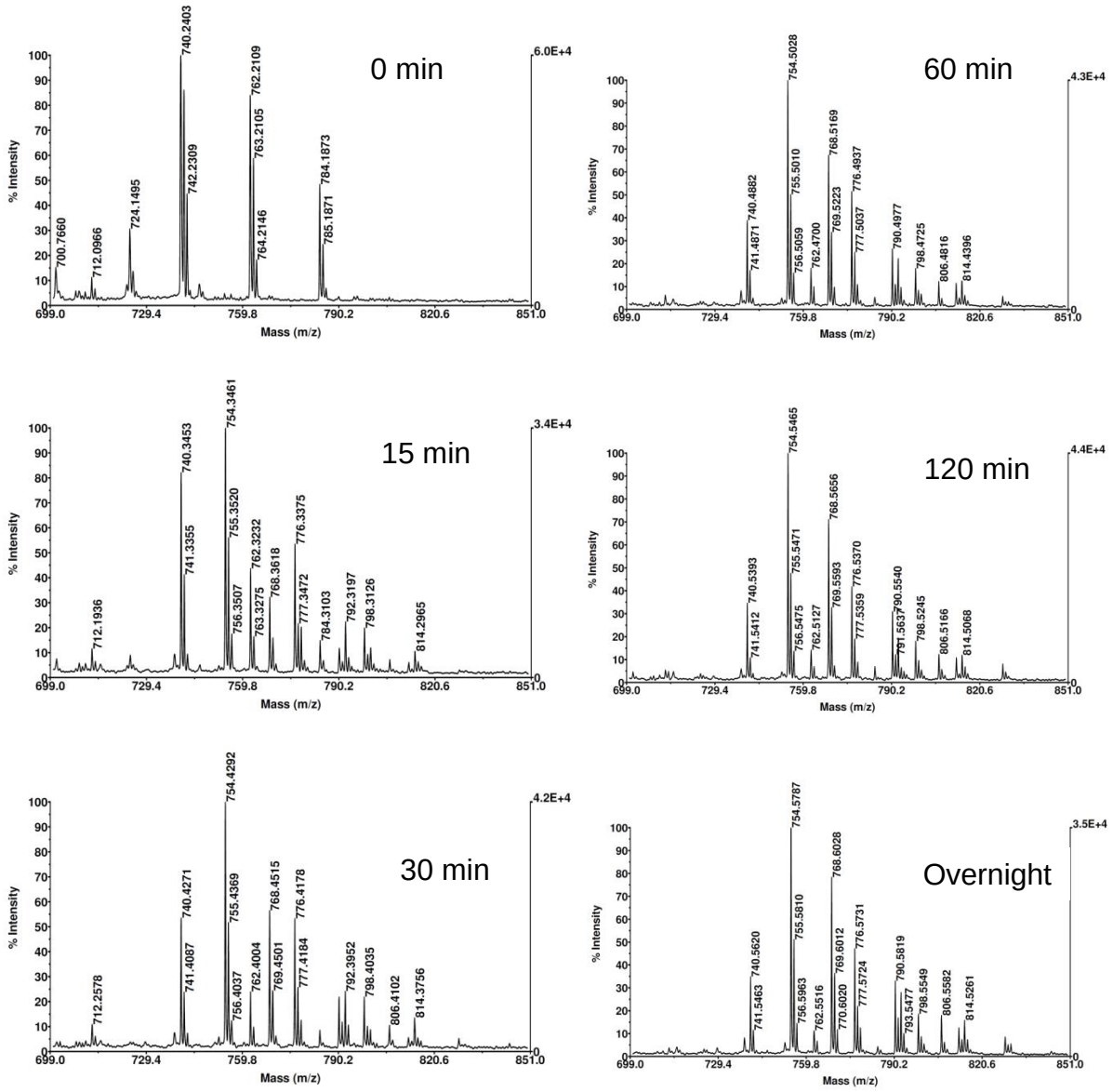


Figure 11. Methylation progression of RPKRIA with NTMT1.

RPKRIA – NTMT2

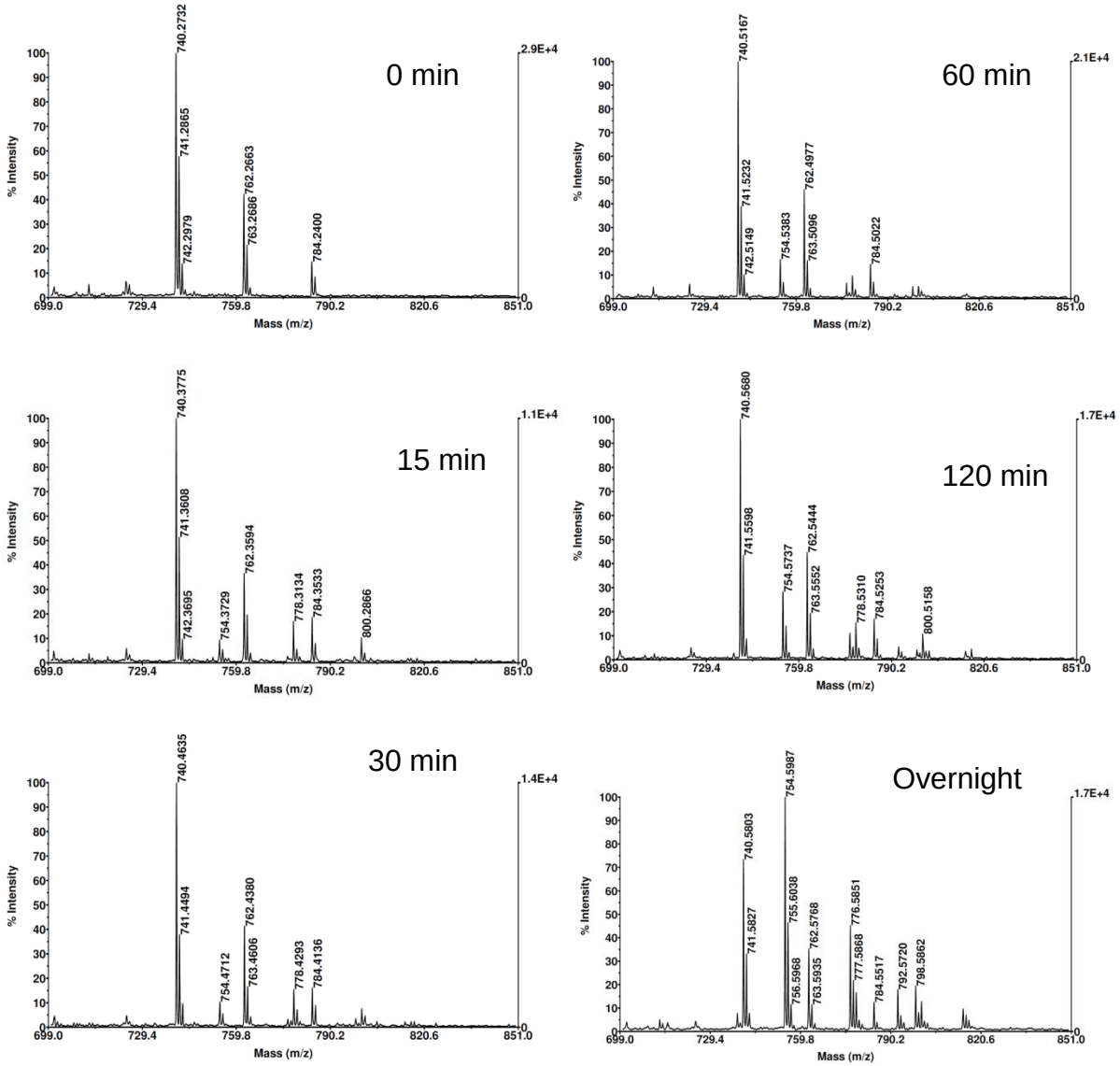


Figure 12. Methylation progression of RPKRIA with NTMT2.

YPKRIA – NTMT1

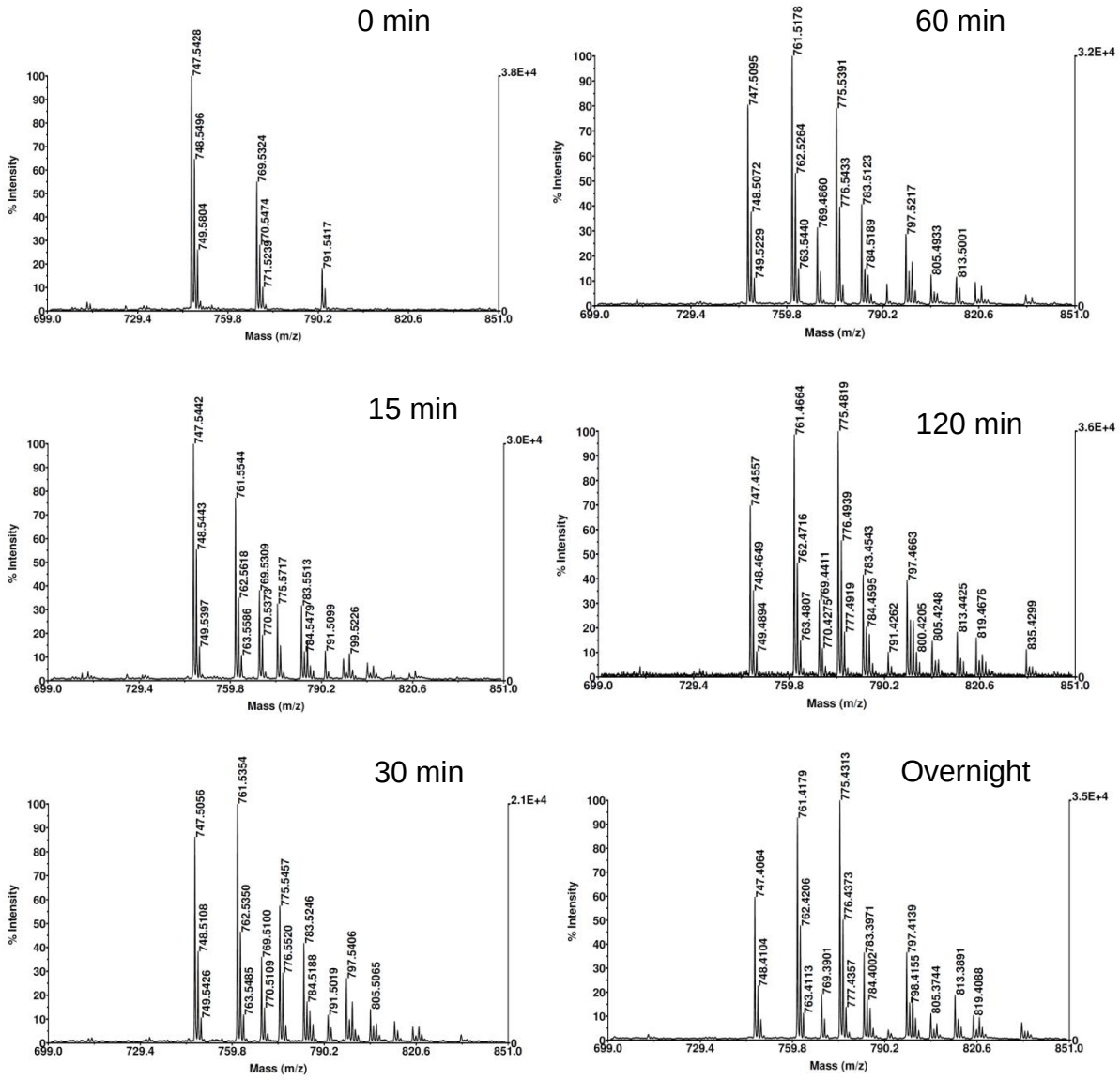


Figure 13. Methylation progression of YPKRIA with NTMT1.

YPKRIA – NTMT2

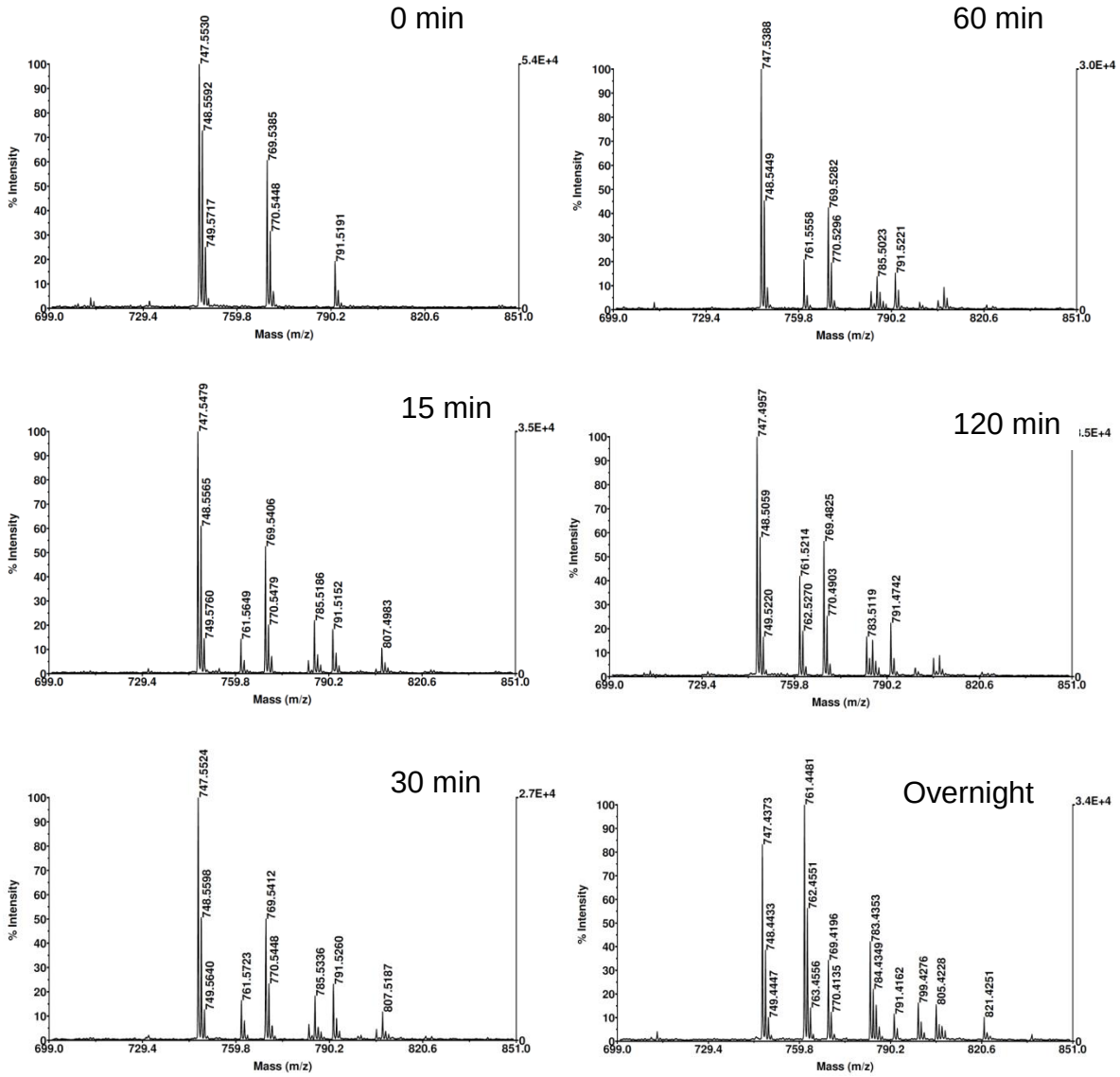


Figure 14. Methylation progression of YPKRIA with NTMT2.

QPKRIA – NTMT1

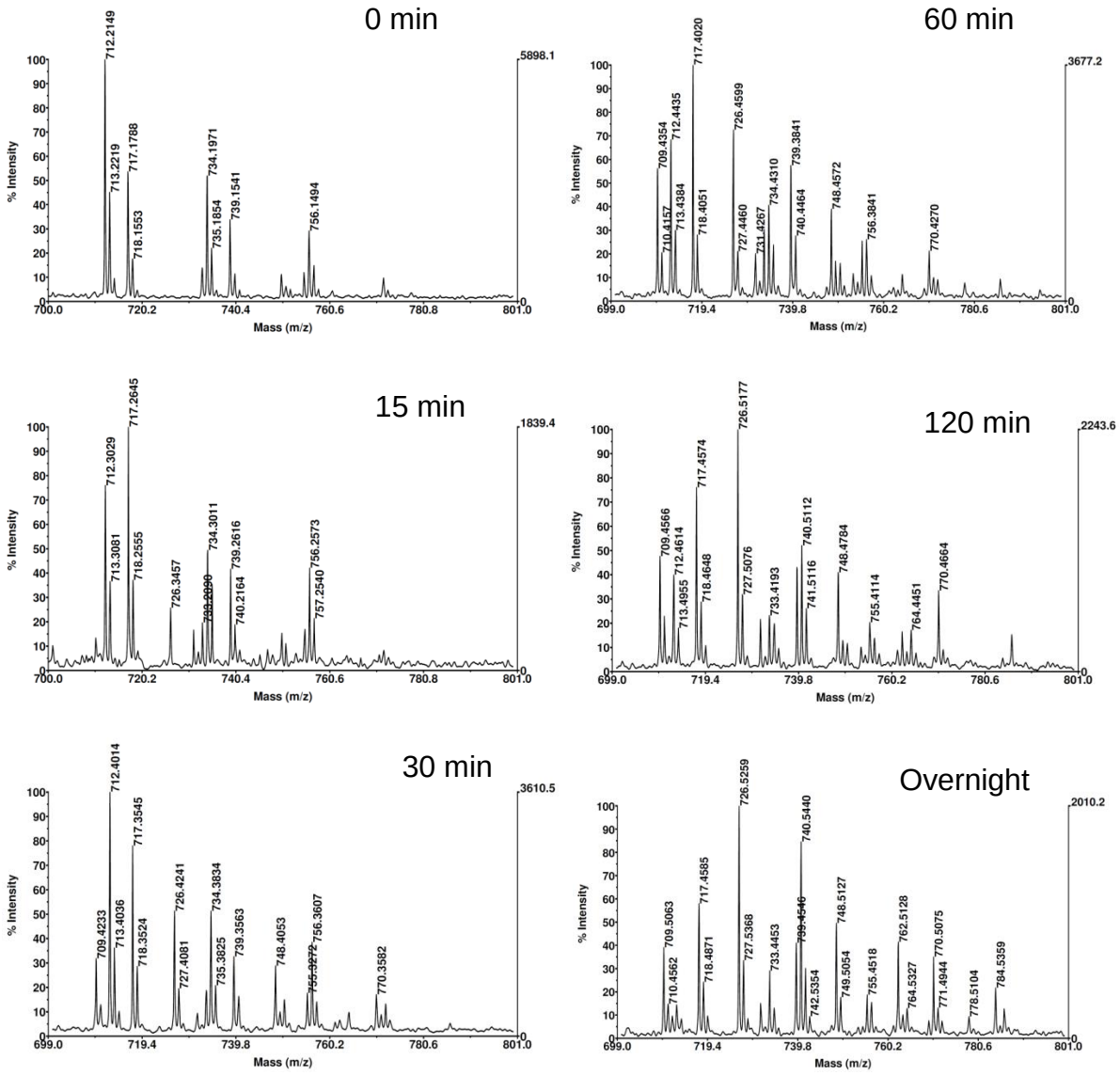


Figure 15. Methylation progression of QPKRIA with NTMT1.

QPKRIA – NTMT2

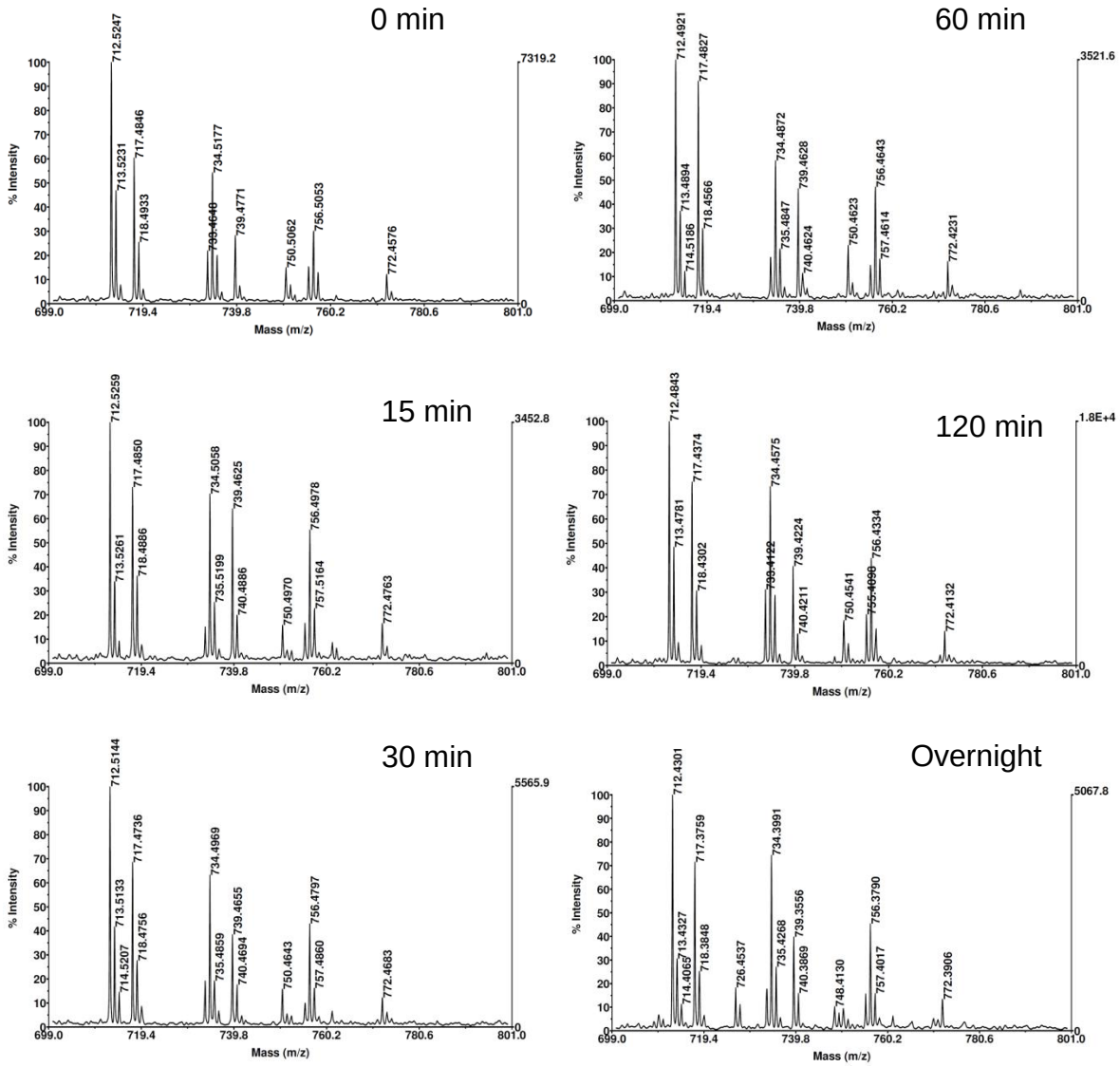


Figure 16. Methylation progression of QPKRIA with NTMT2.

DPKRIA – NTMT1

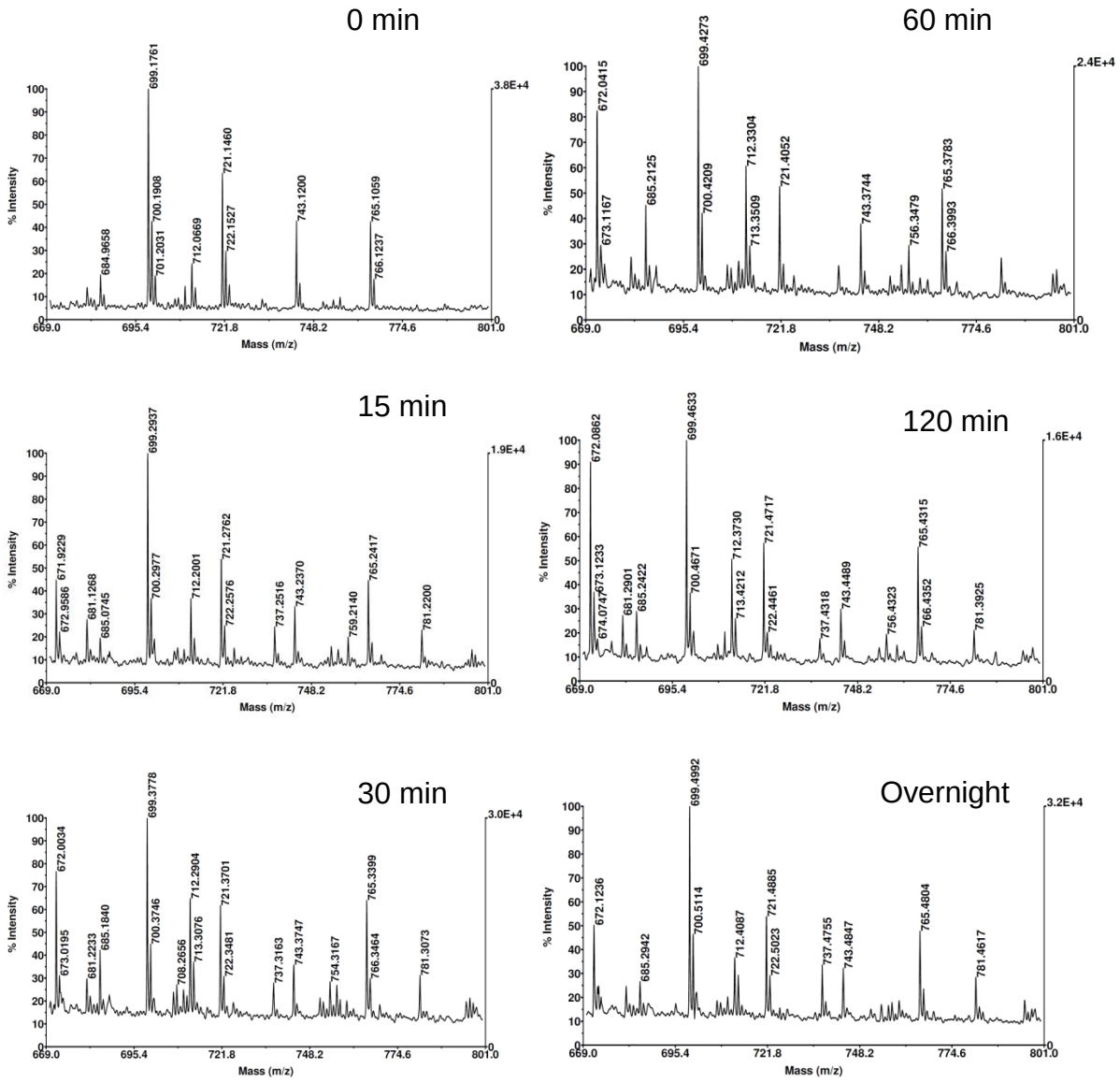


Figure 17. Methylation progression of DPKRIA with NTMT1.

DPKRIA – NTMT2

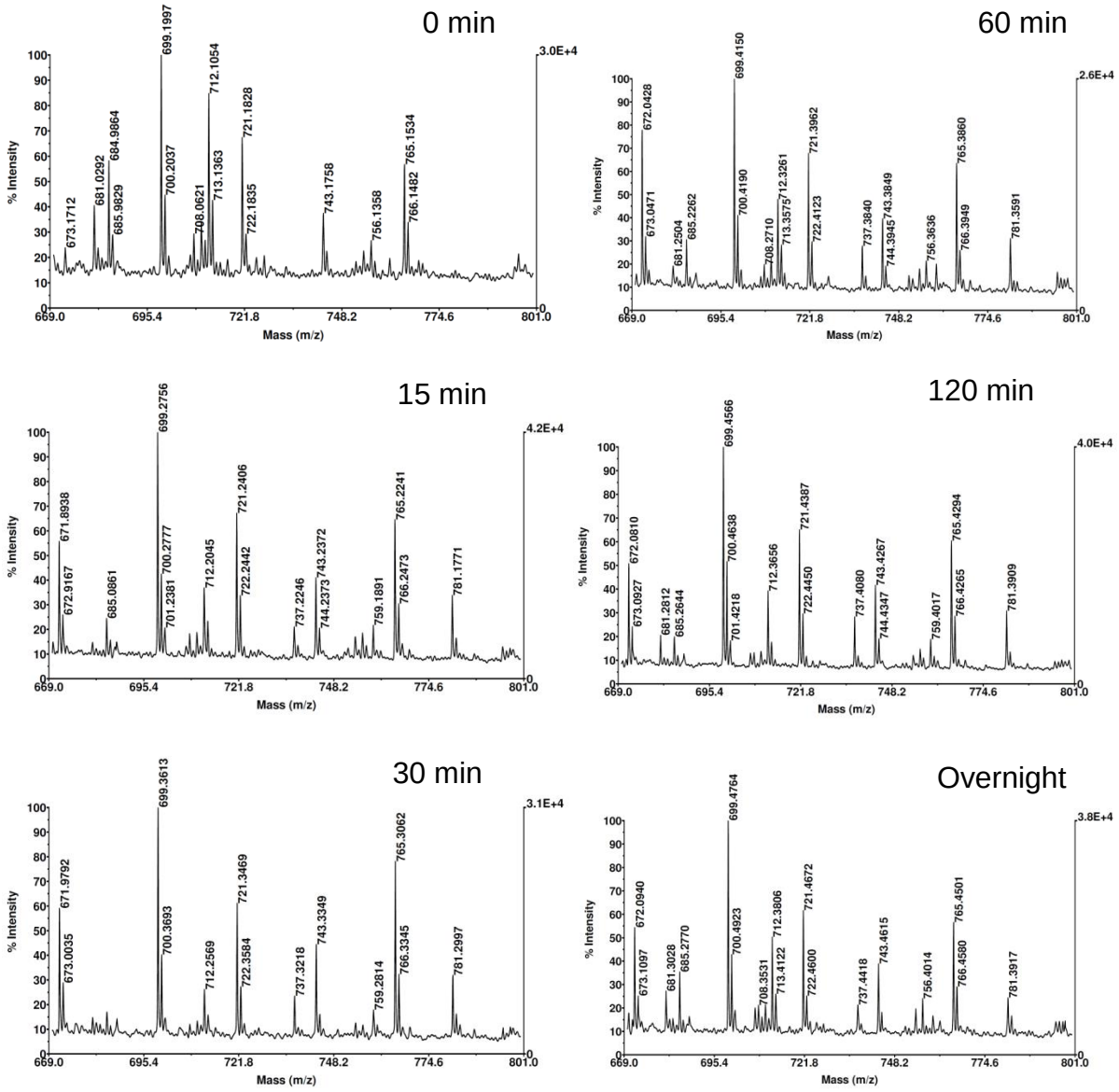


Figure 18. Methylation progression of DPKRIA with NTMT2.

WPKRIA – NTMT1

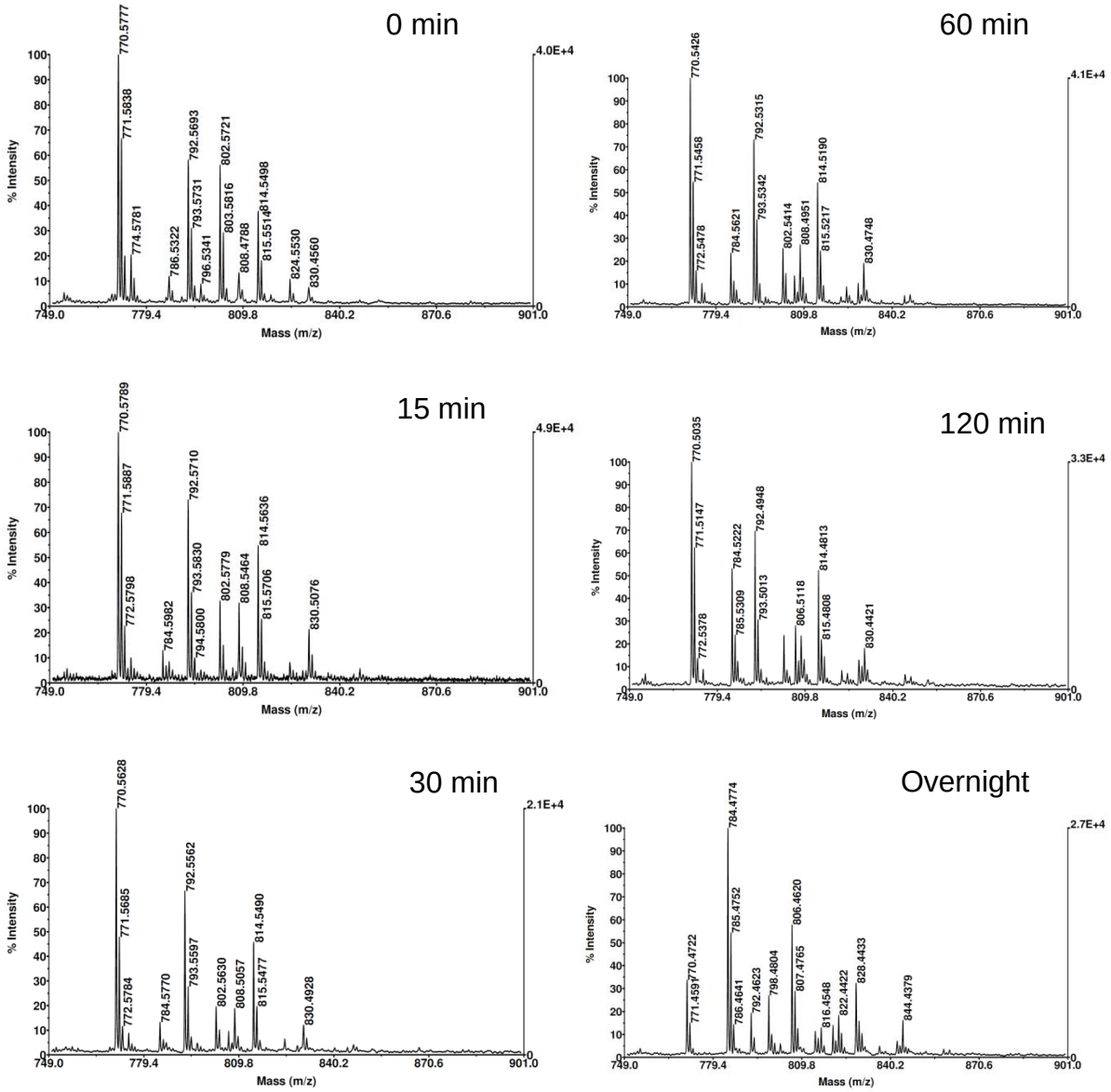


Figure 19. Methylation progression of WPKRIA with NTMT1.

WPKRIA – NTMT2

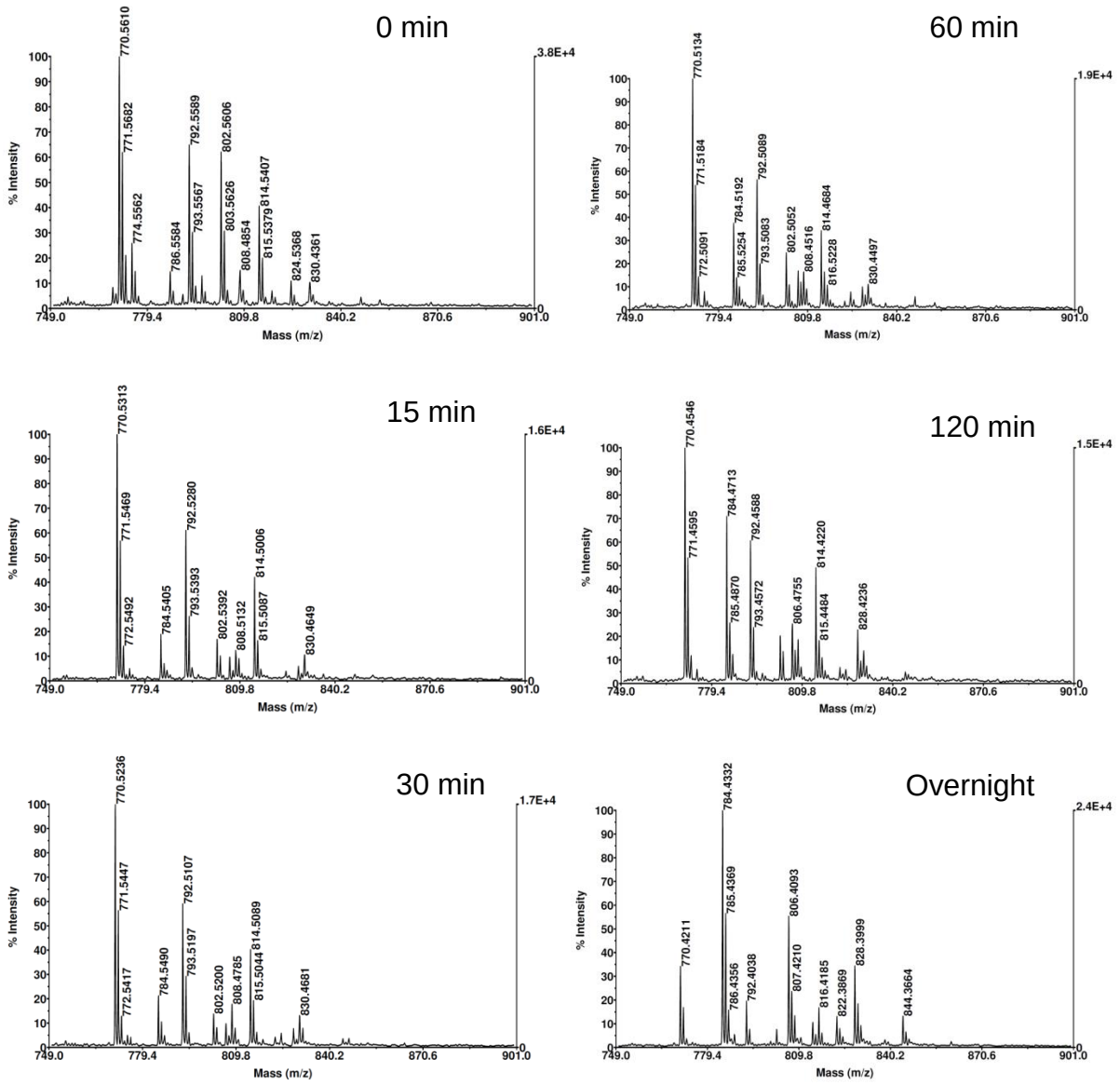


Figure 20. Methylation progression of WPKRIA with NTMT2.

LPKRIA – NTMT1

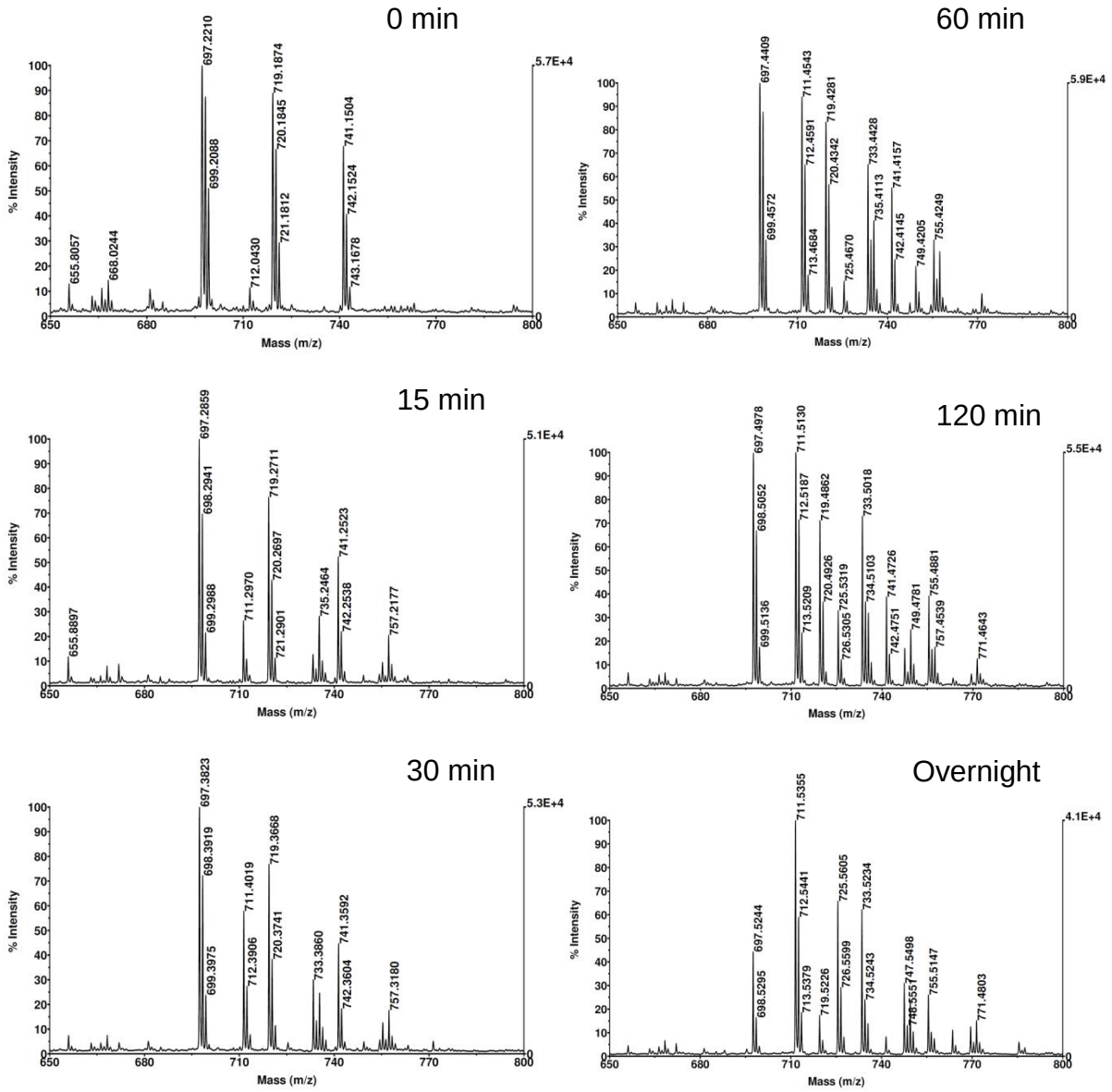


Figure 21. Methylation progression of LPKRIA with NTMT1.

LPKRIA – NTMT2

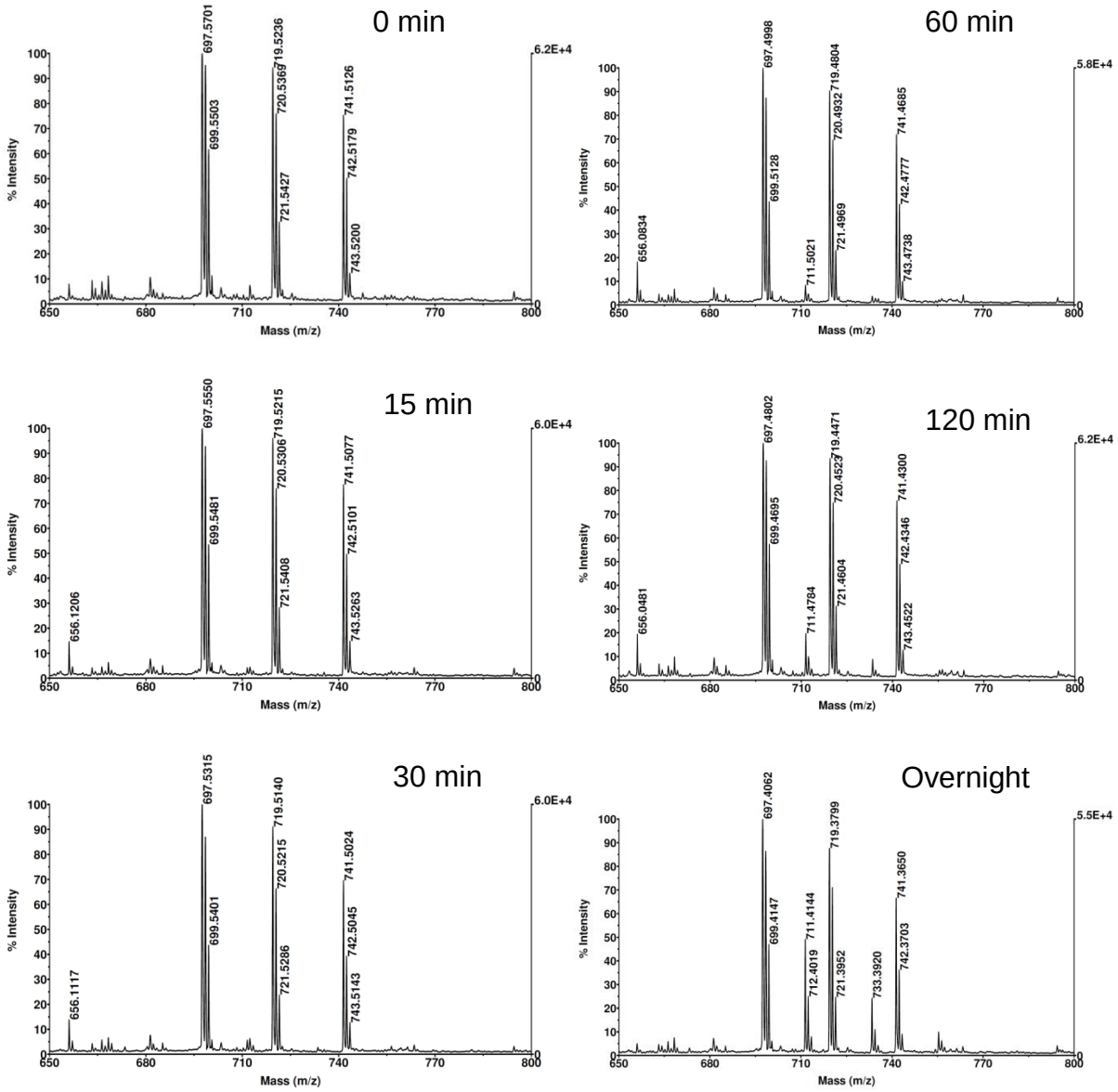


Figure 22. Methylation progression of LPKRIA with NTMT2.

NPKRIA – NTMT1

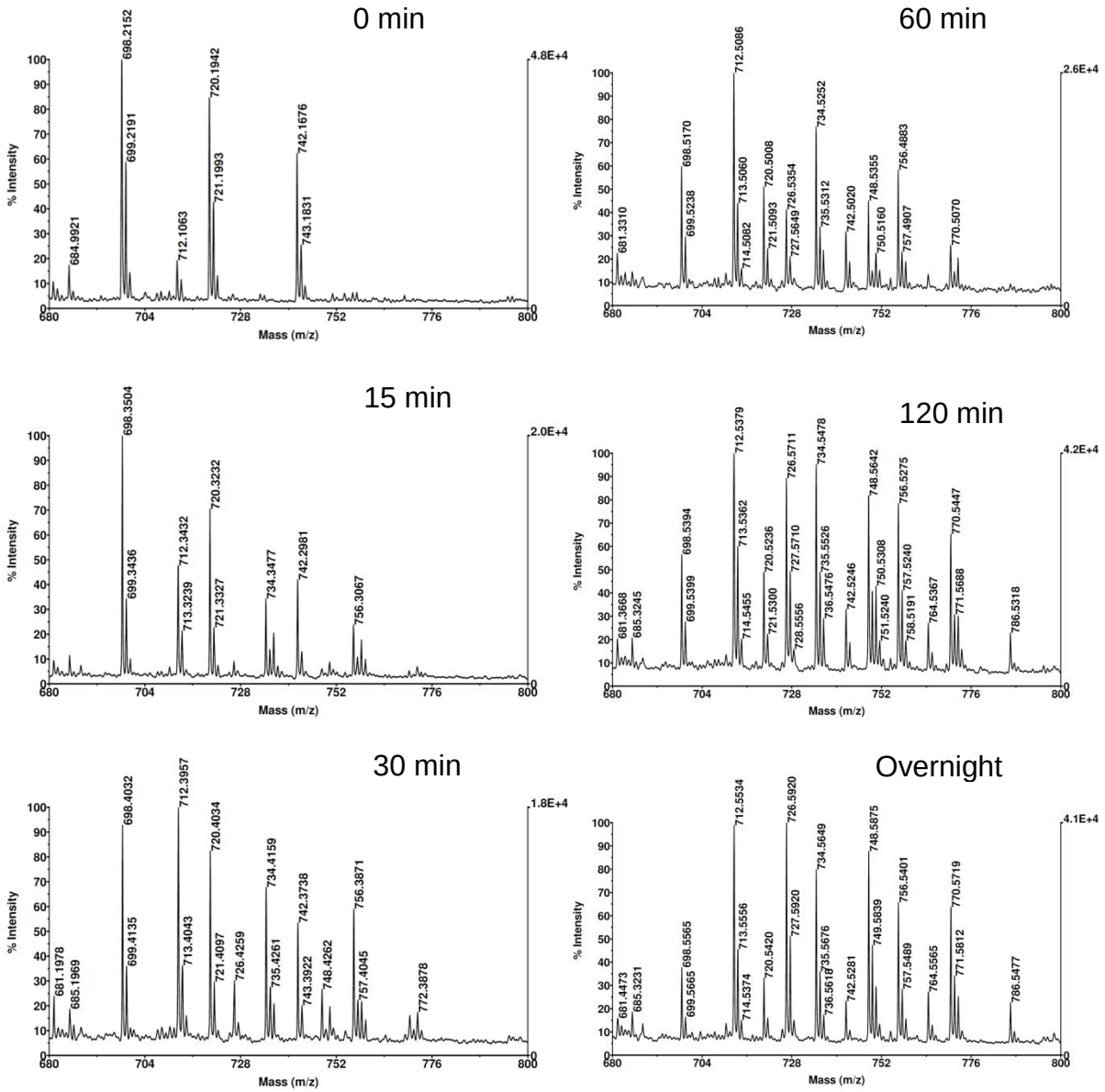


Figure 23. Methylation progression of NPKRIA with NTMT1.

NPKRIA – NTMT2

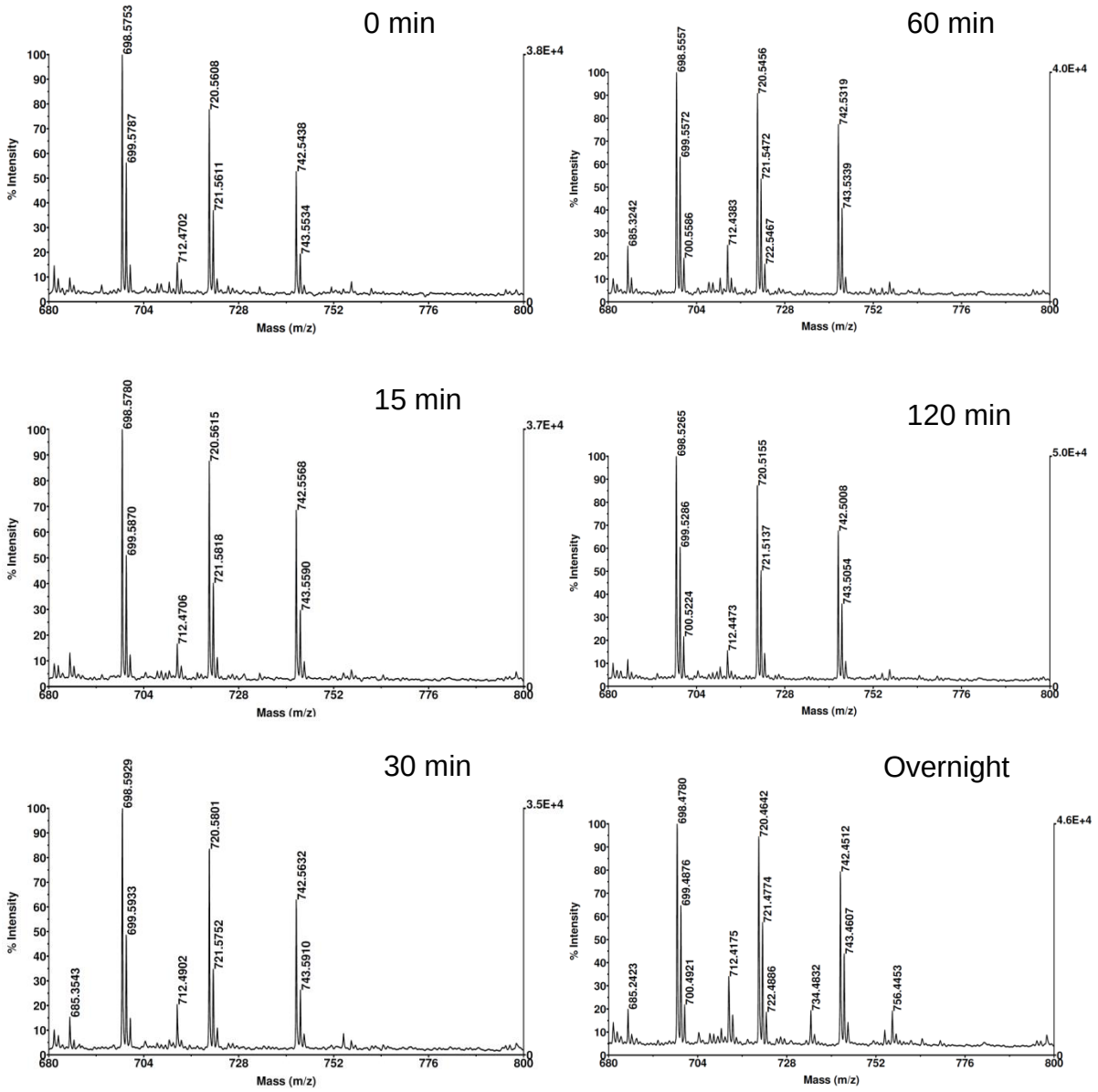


Figure 24. Methylation progression of NPKRIA with NTMT2.

MePPKRIA – NTMT1

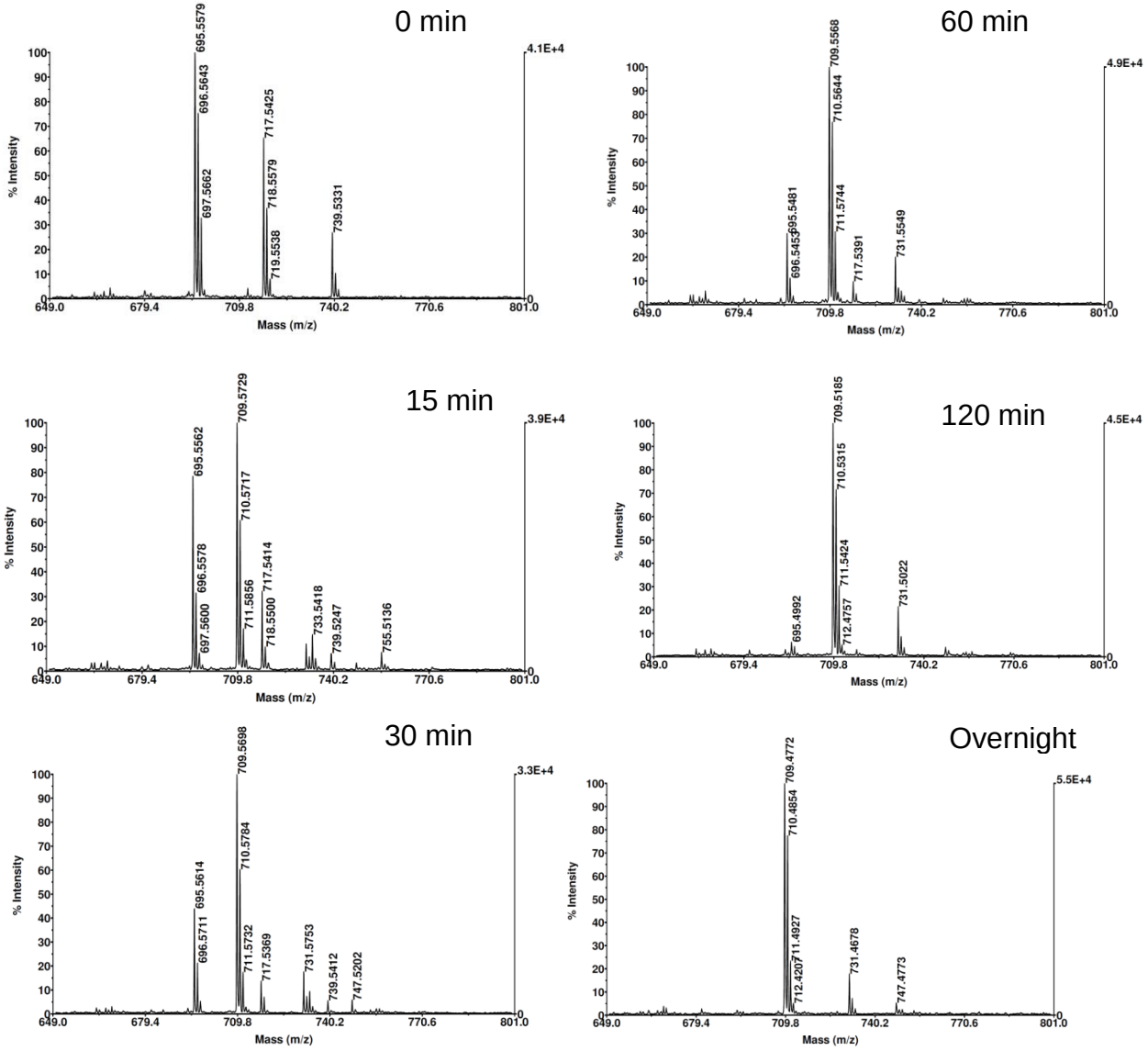


Figure 25. Methylation progression of MePPKRIA with NTMT1.

MePPKRIA – NTMT2

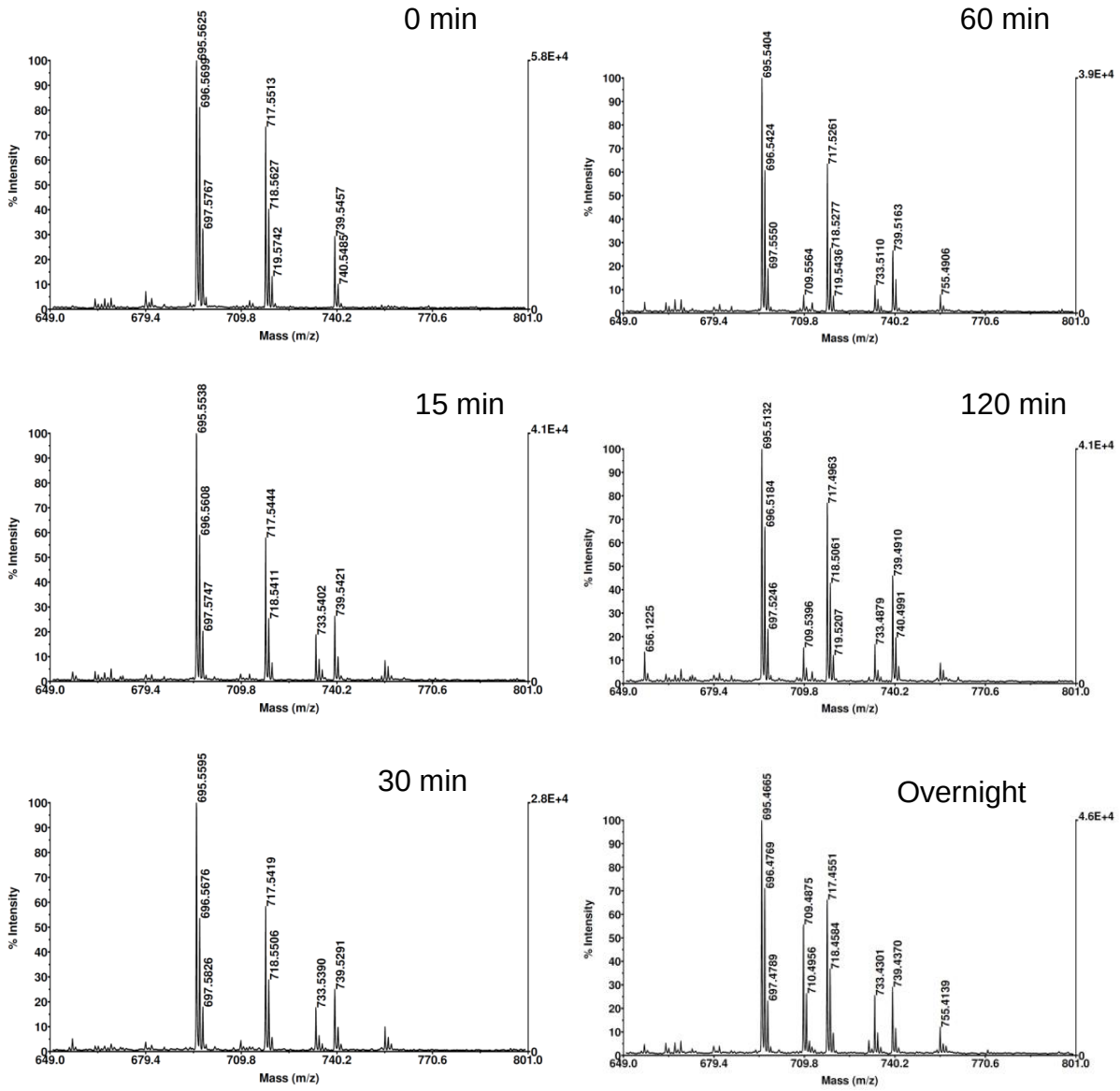


Figure 26. Methylation progression of MePPKRIA with NTMT2.

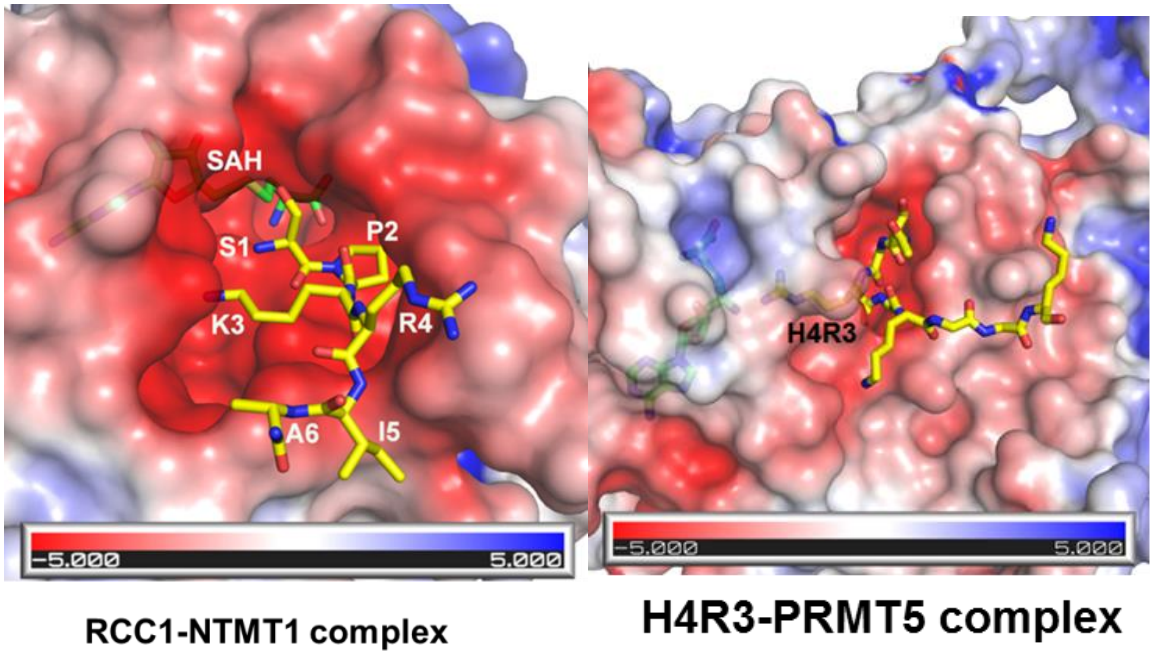


Figure 27. Peptide substrate binding sites of NTMT1 and PRMT5.

Vita

Yunfei Mao was born May 10, 1988 in Sichuan, China. He spent his childhood in Chengdu, China and Hiroshima, Japan. He finished his high school education in Guangya school, Dujiangyan, Sichuan, China. Subsequently, he went to the Waterford Institute of Technology, Waterford, Ireland to pursue his bachelor of science in pharmaceutical science, where he obtained his bachelor's degree in 2012. He joined the Department of Medicinal Chemistry of Virginia Commonwealth University in the same year to pursue his Ph.D.